A variable North Atlantic sink for anthropogenic CO_2: modelling observed change

Submitted by **Alice Lebehot** to the University of Exeter as a thesis for a degree of Doctor of Philosophy in Geography in January 2018.

Supervisors

Dr Paul Halloran, University of Exeter (lead supervisor)

Prof Andrew Watson, University of Exeter

Dr Doug McNeall, UK Met Office

Dr Ute Schuster, University of Exeter

Jury

Prof Laurent Bopp, CNRS, ENS external examiner

Prof Pierre Friedlingstein, University of Exeter internal examiner

Funding

This Ph.D. was supported by NERC and the Met Office through a case-studentship within the RAGNARoCC project.

“The most monumental things we have as humans arise from shared passions. We put ourselves out there and let it happen. The pain, the joy, validation, rejection... Life. If it’s a battle let it be a great one. If there even is an opponent let it be a worthy one. Otherwise why get involved? Great pain, great joy, rejection and validation. Nothing less than that.”

— Patrice Bart-Williams

Abstract

To determine the maximum carbon dioxide (CO_2) emissions consistent with a given global warming threshold, the scientific community must robustly quantify what proportion of human emitted CO_2 will be taken up by the terrestrial and marine carbon reservoirs. The North Atlantic Ocean is a region of intense uptake of atmospheric CO_2. To assess how the North Atlantic CO_2 sink has evolved over the past decades and understand the mechanisms involved in that uptake, observations and models are used. To appreciate the strengths and limitations of observation-based and modelled products, I explore the sources of uncertainties of two widely-used biogeochemical observational products (GLODAP and SOCAT), and carefully evaluate the latest generation of Earth System Models (ESMs) (i.e. the CMIP5 models) against these data. The lack of robust uncertainties on observation-based estimates of the North Atlantic CO_2 uptake has so far limited the community's ability to use observed trends to evaluate CO_2 uptake behaviour simulated by the models. Here, by making use of the strengths of observation-based and modelled products, a novel gap-filling and uncertainty assessment method is developed to (1) robustly quantify the recent change in the basin-wide North Atlantic CO_2 sink, and (2) evaluate simulations of the recent uptake in ESMs. Through the assessment of robust interpolation uncertainties on the annually-varying North Atlantic CO_2 uptake and on the resulting trends over the period 1992-2014, I find that (1) the North Atlantic CO_2 uptake increased at a rate of 0.081 ± 0.012 PgC$\cdot$yr$^{-1}\cdot$decade^{-1} from 1992-2014, corresponding to an additional uptake of 2.2 PgC over this interval relative the flux in the 1992, and (2) state-of-the-art ESMs are consistently biased to lower trend values, with a mean that is about three times smaller than the observation-based trend, equating to an additional uptake of only 0.72 ± 0.40 PgC over the period 1992-2014. I further show that the inability of these models to capture the observed increase in CO_2 uptake is due primarily to biases in modelled ocean biogeochemistry, which I explore through comparison with observations. Our current understanding of the ocean carbon-cycle, as synthesised by ESMs, cannot explain the recent behaviour of the North Atlantic CO_2 sink. Current projections may therefore underestimate the contribution of the North Atlantic to mitigating increasing future atmospheric CO_2 concentrations.

Contents

List of Figures

List of Tables

List of Acronyms

AMOC Atlantic Meridional Overturning Circulation

AR5 Fifth Assessment Report

AS Absolute Salinity

BATS Bermuda Atlantic Time-series Study

CDO Climate Data Operator

CICE Los Alamos National Laboratory sea ice model

CLIVAR Climate and Ocean: Variability, Predictability and Change

CMIP5 Coupled Model Intercomparison Project Phase 5

CRM Certified Reference Material

CT Conductive Temperature

CURB CO^2 Carbon Uptake Revisited - Biases Corrected using Ocean Observations

DIC Dissolved Inorganic Carbon

DIVA Data-Interpolating Variational Analysis

ECCO2 Circulation and Climate of the Ocean Phase II

ECMWF European Centre for Medium-Range Weather Forecasts

ESGF Earth System Grid Federation

ESM Earth System Models

ETOPO Earth Topography

HadOCC Hadley Centre Ocean Carbon Cycle

QC Quality Control

GEOSECS Geochemical Ocean Sections

GLODAP Global Ocean Data Analysis Project

GO-SHIP Global Ocean Ship-based Hydrographic Investigations Program

GSW Gibbs Seawater

IPCC Intergovernmental Panel on Climate Change

LDEO Lamont-Doherty Earth Observatory

LSW Labrador Sea Water

MAE Mean Absolute Error

MLD Mixed Layer Depth

MLR Multi Linear Regression

NAC North Atlantic Current

NAO North Atlantic Oscillation

NCAR National Center for Atmospheric Research

NCEP National Centers for Environmental Prediction

NEMO Nucleus for European Modelling of the Ocean

NPZD Nutrient Phytoplankton Zooplankton Detritus

OCADS Ocean Carbon Data System

OISST Optimum Interpolation Sea Surface Temperature

PFT Phytoplankton Functional Types

RCP Representative Concentration Pathways

RMSE Root Mean Square Error

SOCAT Surface Ocean CO_2 Atlas

SLP Sea Level Pressure

SOP Standard Operation Procedure

SSS Sea Surface Salinity

SST Sea Surface Temperature

TA Total Alkalinity

TTD Transit Time Distribution

VINDTA Versatile INstruments for the Determination of Titration Alkalinity

VOS Voluntary Observing Ships

WOA World Ocean Atlas

WOCE World Ocean Circulation Experiment

Declaration of Authorship

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Chapter 1

Introduction

This chapter aims to present the scientific context and challenges associated with the roles of the oceans in the carbon cycle, which as such, will provide a useful background for the following chapters. Specifically, this chapter will address the questions: What is the role of carbon dioxide (CO_2) in the climate system? How do the oceans contribute into mitigating climate change? Why is the North Atlantic a key region for atmospheric CO_2 uptake? What are the challenges in quantifying that uptake? How would this sink evolve under on-going rising atmospheric CO_2 concentrations? Finally, this chapter will present the subsequent thesis plan.

1.1 CO_2 and the climate system

The Earth's climate system is powered by incoming solar energy and is described by how this solar radiation interacts with the biosphere (e.g. provision of photosynthetically active radiation), hydrosphere (e.g. heating surface ocean), lithosphere (e.g. reflective properties of certain land surfaces) and atmosphere (e.g. latent heat, reflective properties of certain clouds, greenhouse gases) (Figure 1.1; Cubasch et al., 2013). The complex interactions within the climate system lead to a global energy balance between the amount of radiation that goes into the system and the amount that goes back into space (Le Treut et al., 2007). Since the industrial era began in 1750, the intensified interactions between the Earth's spheres and humankind has introduced a "new sphere" to the system, known as the "anthroposphere". The anthropogenic influence on the radiative balance takes place through (1) changes in the

vegetation (i.e. land-use change), which mainly impact the surface albedo and therefore the amount of reflected solar radiation, and (2) the release of particulates and chemicals into the atmosphere (e.g. aerosols, carbon dioxide, methane), which on balance amplifies the greenhouse effect and therefore increases the amount of long-wave radiation that are absorbed by the Earth's surface (Figure 1.1; Cubasch et al., 2013).

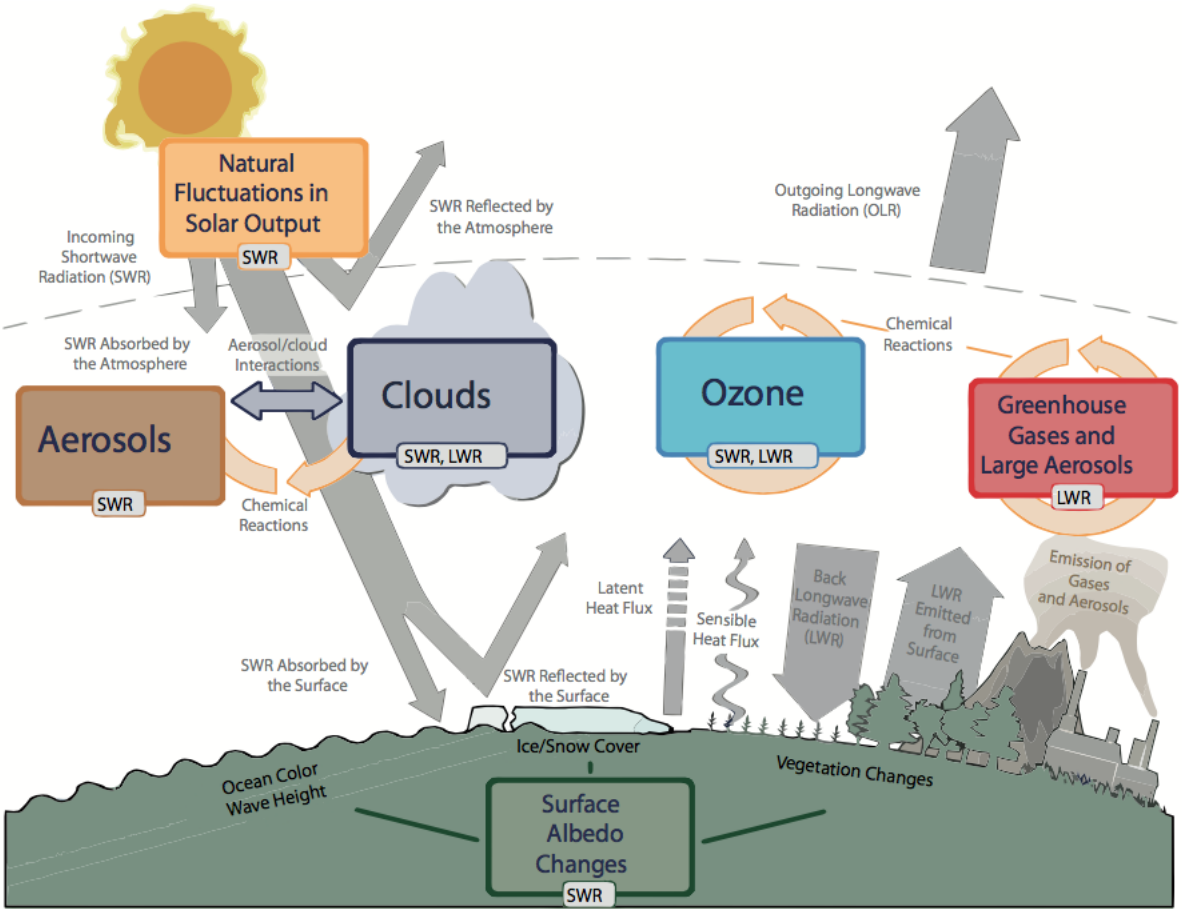

Figure 1.1: The climate system and the dominant drivers of its change. Schematic representation of the radiative interactions within the climate system: from the incoming solar short-wave radiation (SWR) to the outgoing long-wave radiation (LWR). The climate disruptors are indicated by the main boxes. Source: Cubasch et al., 2013

Carbon dioxide (CO_2) is the second most abundant greenhouse gas present in the atmosphere (water vapour being the first). Prior to the industrial era, the levels of atmospheric CO_2 mixing ratio were about 280 parts per million (ppm) (Figure 1.2) and were controlled by natural fluxes within the Earth system (e.g. volcanic eruptions, the difference between the photosynthesis and respiration reactions, where a positive difference would result in the capture of CO_2 by photoautotroph organisms) (Le Treut et al., 2007). Since ~1750, the

atmospheric CO_2 mixing ratio has been increasing, recently reaching levels that are about 44% higher than the preindustrial average (Figure 1.2). The rapid increase in atmospheric CO_2 concentrations is induced by the release of anthropogenic CO_2 emissions into the atmosphere, mostly from the combustion of fossil fuels, the production of cement and the changes in land surface (e.g. deforestation, afforestation) (Figure 1.3a; Le Quéré et al., 2015). During the period 1959-2014, approximately 324 PgC and 75 PgC were emitted from fossil fuels and land-use change activities, respectively. Nevertheless, about 56% of the total emissions (i.e. the sum of fossil fuel and land-use change emissions) emitted each year is removed from the atmosphere through the absorption of carbon by the natural terrestrial and marine sinks (Le Quéré et al., 2015). The efficiency of those carbon sinks is described by the airborne fraction, which is the ratio between the atmospheric growth rate and the total emissions (Figure 1.3b). A decrease in the airborne fraction indicates a strengthening in the natural carbon sinks, as observed since the 2000s (Ballantyne et al., 2012; Keenan et al., 2016). Nevertheless, the detection of significant trends in the highly variable airborne fraction, that is associated with various sources of uncertainties (Ballantyne et al., 2015), and therefore in the sinks' efficiency can be challenging (e.g. Canadell et al., 2007; Ballantyne et al., 2012). The high inter-annual variability in the airborne fraction arises from the response of the sinks (mostly the land sink) to natural climate variability, as well as from the impacts on the carbon cycle (Keenan et al., 2016). It is therefore crucial to determine with confidence the terrestrial and marine carbon sinks' response to natural variability and climate change in order to quantify the maximum threshold of CO_2 emissions per degree of warming (Le Quéré et al., 2015).

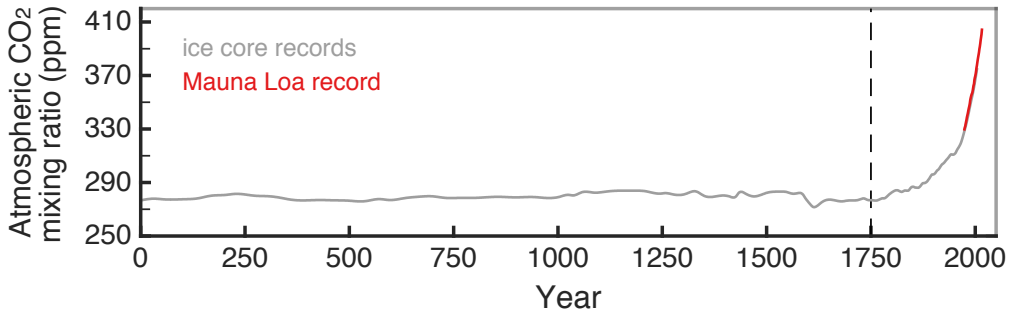

Figure 1.2: Variations in atmospheric CO_2 mixing ratio. CO_2 mixing ratio measured from ice cores (gray) (Etheridge et al., 1996; Etheridge et al., 1998; MacFarling Meure et al., 2006; Etheridge, 2010) and from the Mauna Loa station in Hawaii (red), which is the longest running observed archive of atmospheric CO_2 concentrations (Thoning et al., 2016; Le Quéré et al., 2015).

During the period 1959-2010, the net global carbon uptake by the land and oceans has increased in response to the rising atmospheric CO_2 (Ballantyne et al., 2012; Ballantyne et al., 2015). The uptake between the two reservoirs is similar, with about 30% of the total emissions being absorbed by the land and 26% by the oceans (Figure 1.3c,d) (Le Quéré et al., 2015). The quantification of the intensity and variability of those sinks at both the global and regional scale is however challenging, as it relies on (1) localised observations in a heterogenous terrain, which require interpolation methods that may underestimate the unknown true spatial and temporal variability as statistical methods are likely to not represent extreme values, if even sampled in the first place, and (2) on models that are, by definition, simplified numerical versions of the real-world (Le Quéré et al., 2015; Ballantyne et al., 2015). For instance, the global carbon budget, whose results are widely relayed to stakeholders, estimates the global marine CO_2 sink (mean, anomalies and trend) prior to the 1990s solely from biogeochemical ocean models forced with meteorological reanalysis data, and after the 1990s from the combination of those models with two observational-interpolated products (Le Quéré et al., 2015). As such, without the quantification of robust uncertainties on interpolated products, as well as a clear understanding on the impacts of potentially important missing mechanisms or parametrisation choices in the models, it is yet difficult to fully determine the variability of the ocean sink and hence detect CO_2 oceanic saturation stages (Ballantyne et al., 2015). The land sink is also particularly difficult to quantify, due to (1) the relatively low carbon mobility across the vegetation/soil (as opposed to the oceans which - to some extent - homogenise their carbon content through mixing with the surrounding waters), (2) the large variability in the magnitude of the air-land CO_2 fluxes (large negative and positive fluxes), (3) the high spatial variability of vegetation types, (4) the wide range of vegetation/soil responses to climate change conditions, and (5) the challenging boundary distinctions with land-use change activities (Le Quéré et al., 2015). Since those processes are still poorly known, mostly due to the lack of robust measures of territorial carbon stocks across large spatial areas, and therefore poorly represented in vegetation models, the land sink is commonly estimated as the residual from the other terms of the global carbon budget (i.e. total emissions, atmospheric CO_2 growth and ocean sink). By closing the carbon budget, the land sink estimate also contains the various sources of errors and uncertainties from those

other terms (Ballantyne et al., 2015), leading to further challenges when understanding the efficiency of the terrestrial sink and evaluating vegetation models. Within this context, it is crucial to provide robust estimates of the carbon budget terms, particularly on the ocean sink.

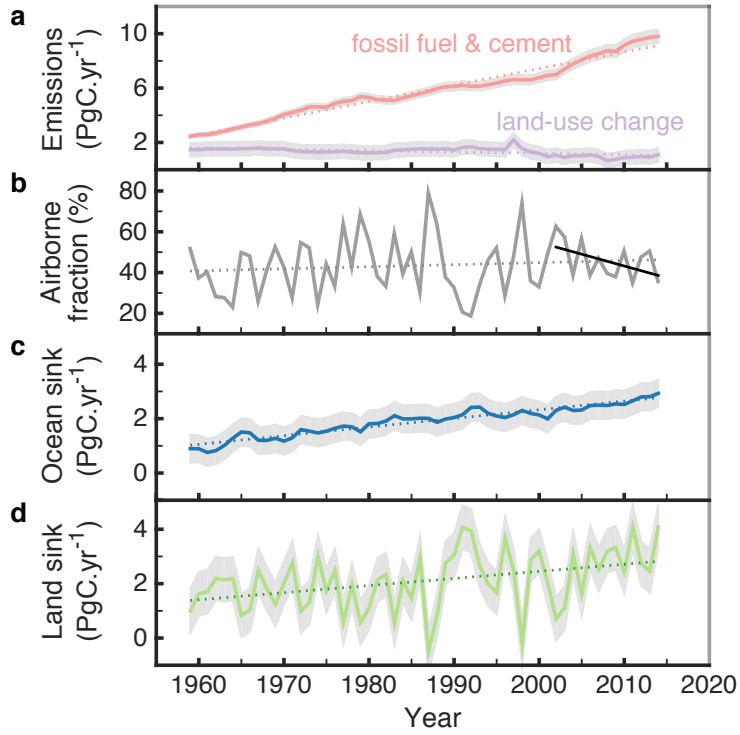

Figure 1.3: The global carbon sources and sinks. **a**, Total emissions of carbon from fossil fuel combustion, cement production (red) and land-use change (purple) (Boden et al., 2013). **b**, Airborne fraction (AF) was deduced from the atmospheric CO_2 growth rate (Houghton et al., 2012) and the total carbon emission (Boden et al., 2013). **c**, Ocean and **d**, land carbon sinks (Le Quéré et al., 2015). Note that the spike in the land sink in 1997 is due the addition of inter-annual variability information given by satellite-based measurements (Le Quéré et al., 2015). Dotted lines represent the linear fit of each time series over 1959-2014 (trend in the AF is statistically insignificant over this interval, at the 5% significance level). Dark line in **b** is the linear fit from 2002, with a p-value of 0.09. Grey shadings correspond to the uncertainties described in Le Quéré et al., 2015.

1.2 CO_2 in seawater

By absorbing about a third of the human-emitted CO_2 over the industrial period, the oceans play a fundamental role in controlling the levels of atmospheric CO_2 and hence mitigating climate change (Khatiwala et al., 2013; Le Quéré et al., 2015). The repartition of the anthro-

pogenic carbon uptake across the world's oceans is however not uniform. The North Atlantic Ocean has the highest column inventory of anthropogenic carbon, while other regions present relatively low inventories (Figure 1.4). The spatial distribution and the underlying temporal variability of the anthropogenic carbon uptake, which is occurring on top of the natural carbon cycle, involve various biogeochemical processes that are intrinsically linked to the ocean circulation and atmospheric conditions. This section therefore aims to shed light on the main chemical (Section 1.2.1) and biological (Section 1.2.2) processes and concepts of the marine carbon cycle.

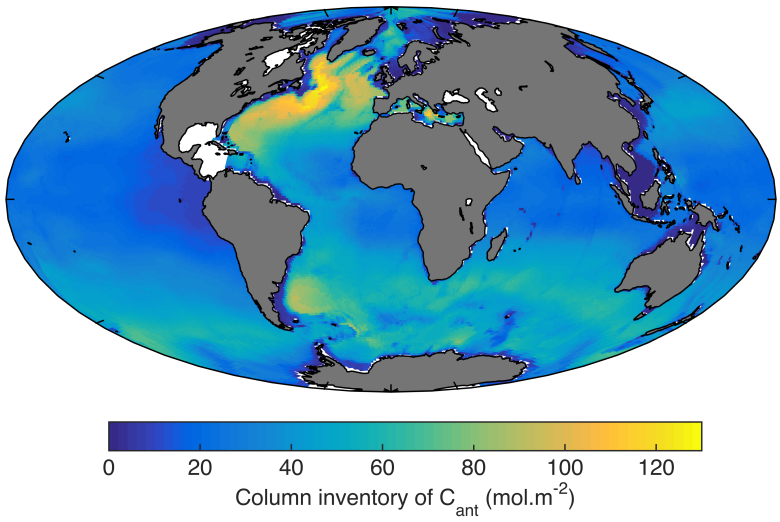

Figure 1.4: Column inventory of anthropogenic carbon. The water column inventory was calculated using the climatology of anthropogenic carbon concentrations representative of 2002 from Lauvset et al., 2016.

1.2.1 Inorganic chemical processes

The marine carbon chemistry is described by four key variables: partial pressure of CO_2 in seawater ($pCO_{2\text{-ocean}}$), DIC, TA and pH (Millero et al., 2002). To bring a relevant background to the rest of the thesis, this section focuses on introducing the fugacity of CO_2 (which is similar to $pCO_{2\text{-ocean}}$ and linked the air-sea CO_2 flux by partially influencing the air-sea CO_2 gradient), surface DIC and TA, and discusses how these variables condition the amount of additional CO_2 that can be taken-up by the oceans.

Fugacity

The fugacity of CO_2 in both the ocean ($fCO_{2\text{-ocean}}$) and the atmosphere ($fCO_{2\text{-atmosphere}}$) are key variables for controlling the direction of the air-sea CO_2 flux. Since the fugacity of CO_2 in the atmosphere (which is related to the atmospheric CO_2 mixing ratio; Equation 1.2) is relatively homogenous across the globe, regions of ocean uptake and outgassing are therefore mainly driven by variations in the surface ocean $fCO_{2\text{-ocean}}$ (Sarmiento and Gruber, 2006).

Fugacity is similar[1] to partial pressure, except that it takes into account the non-ideal nature of CO_2 (Equation 1.1). The partial pressure of CO_2 in seawater describes the thermodynamic equilibrium between CO_2 in the gas phase and the concentration in the underlying seawater that is proportional to the solubility of CO_2 in seawater (Sarmiento and Gruber, 2006). Solubility is predominantly temperature-dependent (it also depends on salinity but to a lesser degree) and increases as temperature decreases (and vice-versa). As such, the spatial distribution of $pCO_{2\text{-ocean}}$ (and hence $fCO_{2\text{-ocean}}$) is highly linked to patterns in surface temperatures, with relatively high $pCO_{2\text{-ocean}}$ in equatorial regions and low at high latitudes (Takahashi et al., 2009).

$$fCO_2 = pCO_2 \times \exp\left(P\frac{B+2\delta}{RT}\right) \tag{1.1}$$

$$pCO_{2-\text{atmosphere}} = xCO_{2-\text{atmosphere}} \times 10^{-6} \times (P - pH_2O) \tag{1.2}$$

$$pH_2O = 0.981 \times \exp(14.32602 - 5306.03/T) \times 101325 \tag{1.3}$$

where, P is the sea level pressure (Pa), R the ideal gas constant (8.314 $J{\cdot}K^{-1}{\cdot}mol^{-1}$), T the temperature (K), and B and δ are the temperature dependent coefficients ($m^3{\cdot}mol^{-1}$) (Weiss, 1974; Kortzinger, 1999), $xCO_{2\text{-atmosphere}}$ the CO_2 mixing ratio and pH_{20} the partial pressure of saturated water vapour (Pa) (Cooper et al., 1998).

[1]There is a difference between the two variables of about 1 µatm for a range of temperature between 0°C and 30°C and with a total pressure of 1 atm, which corresponds to ∼0.3% of the mean $fCO_{2\text{-ocean}}$ (Zeebe and Wolf-Gladrow, 2001).

Air-sea CO_2 flux

The direction of the air-sea CO_2 flux (F) is set by the disequilibrium between the fugacities of CO_2 in the atmosphere and surface ocean (Liss and Merlivat, 1986):

$$F = k \times \alpha \times \Delta fCO_2 \tag{1.4}$$

$$\Delta fCO_2 = fCO_{2-\text{atmosphere}} - fCO_{2-\text{ocean}} \tag{1.5}$$

where k is the gas transfer velocity (cm·h^{-1}) and α the solubility (μmol·kg^{-1} · μatm^{-1}). Seawater with a relatively low fugacity leads to a positive disequilibrium and hence a net flux into seawater. The air-sea CO_2 flux therefore presents gradients that follow surface $fCO_{2\text{-ocean}}$ gradients, which themselves are linked to temperature gradients and different ocean circulation regimes. As such, the climatological air-sea CO_2 flux is characterised by regions of outgassing in the tropical belt and uptake in the high latitudes (Figure 1.5; Takahashi et al., 2009).

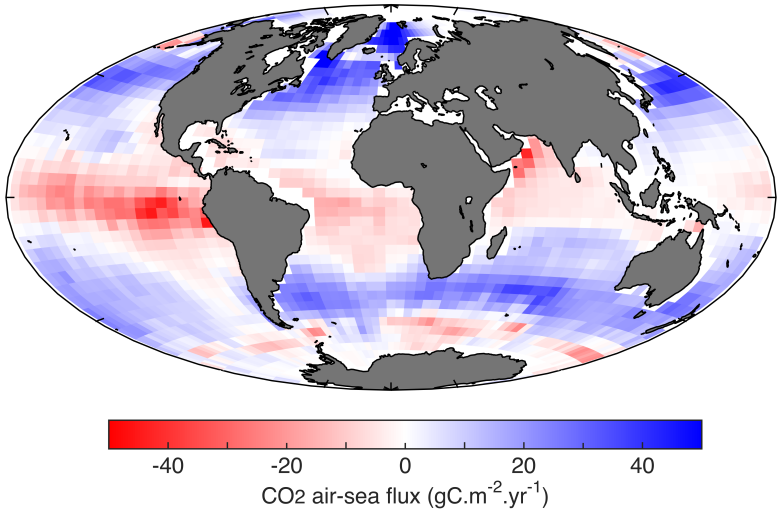

Figure 1.5: Air-sea CO_2 flux climatology. Red indicates CO_2 outgassing region (negative air-to-sea flux) and blue of uptake (positive air-to-sea flux) (Takahashi et al., 2009).

While the direction of the air-sea CO_2 flux is rather easily-constrained at places where surface ocean observations are made, the amplitude of the flux (Equation 1.4) is slightly more challenging to determine (Wanninkhof, 1992), particularly due to the uncertainties associated with the gas transfer velocity parameterisation. The gas transfer velocity aims at characteris-

ing the various physical conditions at the sea surface (e.g. breaking waves, bubbles, sea-ice) and their impact on the CO_2 flux (Couldrey et al., 2016). While the gas transfer velocity is commonly parametrised as a polynomial function of the wind speed (e.g. Nightingale et al., 2000; Takahashi et al., 2009; Wanninkhof, 2014), the definition of this parametrisation and the choice of the wind speed product can lead to a 15% difference on the global air-sea CO_2 fluxes between the methods used (Wanninkhof et al., 2013). Nevertheless, neither the gas transfer velocity or the choice of the parametrisation seems to control the variability of the air-sea CO_2 flux on timescales greater than 5-year long, while ΔfCO_2 does (Couldrey et al., 2016). When studying the variability of the CO_2 flux on pentadal to multidecadal timescales, it is therefore crucial to reduce the uncertainties on ΔfCO_2 and hence on $fCO_{2\text{-ocean}}$, in order to provide robust flux estimates.

Dissolved Inorganic Carbon

The dissolution of gaseous CO_2 into seawater leads to chemical reactions that are fast enough to allow the assumption of thermodynamic equilibrium (Sarmiento and Gruber, 2006). The dissolved gas molecule is first hydrated, forming aqueous CO_2 (CO_2(aq)), whose solubility is highly temperature-dependent (Equation 1.6). The aqueous CO_2 spontaneously reacts with water to form the unstable chemical of carbonic acid (H_2CO_3), which remains at a low concentration (Equation 1.7; Sarmiento and Gruber, 2006). The difficult differentiation between CO_2(aq) and H_2CO_3 is dealt with by the formation of a hypothetical species ($CO_2{}^*$; Equation 1.8), expressed as the sum of both (i.e. CO_2(aq) and H_2CO_3) concentrations (Sarmiento and Gruber, 2006). The reaction of $CO_2{}^*$ with water leads to the formation bicarbonate ($HCO_3{}^-$; Equation 1.9) and carbonate ions ($CO_3{}^{2-}$; Equation 1.10). The dissolution of CO_2 therefore releases protons H^+, identifying CO_2 as an acid, and $HCO_3{}^-$ and $CO_3{}^{2-}$ as the conjugate bases (Sarmiento and Gruber, 2006). Each chemical reaction is temperature, salinity and pressure dependent. As such, the concentration of the individual carbonate species (i.e. $CO_2{}^*$, $HCO_3{}^-$ and $CO_3{}^{2-}$) is not conservative[2]. To describe the carbonate system through conservative prop-

[2]A quantity Q is qualified as conservative when it obeys to a linear mixing relationship and when it remains constant under changes of pressure and temperature (Wolf-Gladrow et al., 2007).

erties, which are particularly useful for modelling exercises, DIC is introduced.

$$\mathrm{CO_2(gas) \rightleftharpoons CO_2(aq)} \tag{1.6}$$

$$\mathrm{CO_2(aq) + H_2O \rightleftharpoons H_2CO_3} \tag{1.7}$$

$$\mathrm{[CO_2(gas) \rightleftharpoons {CO_2}^*]} \tag{1.8}$$

$$\mathrm{{CO_2}^* + H_2O \rightleftharpoons {HCO_3}^- + H^+} \tag{1.9}$$

$$\mathrm{{HCO_3}^- + H^+ \rightleftharpoons {CO_3}^{2-} + 2\,H^+} \tag{1.10}$$

DIC is the sum of the carbonate species' concentrations (Equation 1.11; Williams and Follows, 2011). For global mean surface seawater properties, the speciation of DIC consists of 88.6% $\mathrm{HCO_3^-}$, 10.9% $\mathrm{CO_3^{2-}}$ and 0.5% $\mathrm{CO_2^*}$ (Sarmiento and Gruber, 2006).

$$\mathrm{DIC}=[\mathrm{{CO_2}^*}] + [\mathrm{{HCO_3}^-}] + [\mathrm{{CO_3}^{2-}}] \tag{1.11}$$

As previously explained, the air-sea $\mathrm{CO_2}$ flux is stimulated in cold waters due to the solubility effect. At a fixed $f\mathrm{CO_{2\text{-atmosphere}}}$, cold waters therefore hold more DIC than warm waters (Williams and Follows, 2011), which explains why the spatial distribution of surface DIC is closely related to the distribution of surface temperature[3] (Figure 1.6a). Nevertheless, the surface DIC does not follow a clear meridional gradient, from low concentrations at the equator to high at the poles, indicating that other factors than solubility affect surface DIC. For example, the relatively high DIC concentration in the subtropical gyres could be explained by the strong stratification that inhibits the dilution of the accumulated carbon, and in subpolar gyres by the deepening of the mixed layer during deep water formation events or seasonal changes in the stratification, which entrains DIC-enriched waters to the surface. In the Southern Ocean, DIC-enriched deep waters are also brought up to the surface due to the very active Ekman upwelling generated by the strong westerly winds (Toggweiler et al., 2006).

[3]In the open ocean, salinity has a much smaller impact on DIC than temperature, with a variation of 20 $\mu\mathrm{mol\cdot kg^{-1}}$ for salinities varying from 32.5 to 37.5 psu, corresponding to only 8% of DIC variation from temperature change (Williams and Follows, 2011).

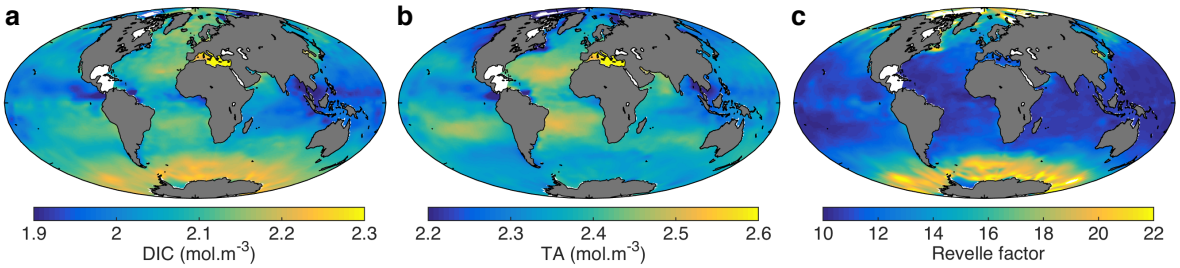

Figure 1.6: Surface climatologies of biogeochemical variables Climatology of **a**, DIC normalised to the year 2002, **b**, TA representing the period 1972-2013 (Lauvset et al., 2016), **c**, Revelle factor calculated using the DIC and TA fields (Equation 1.14). For the purpose of this chapter, errors from the interpolation method (Lauvset et al., 2016) and from the non-ideal steps for calculating the Revelle factor are neglected.

Total Alkalinity

While DIC relies on the conservation of mass, Total Alkalinity (TA), which is also a conservative variable, relates to the conservation of charge. Specifically, TA corresponds to the excess of proton acceptors over proton donors and equals the concentration of acid (i.e. protons (H^+)), that is necessary to neutralise all proton acceptors present in seawater (Wolf-Gladrow et al., 2007). Proton acceptors are distinguished from proton donors depending on the reference species, which is determined from the zero level of protons (Wolf-Gladrow et al., 2007). By convention, the zero level of protons defines the dominant species of an acid-base system at pH=4.5, from which are deduced the number of protons required by the other species of the system to reach the mass balance for hydrogen ion (e.g. if a species must release a proton (H^+) to reach the dominant species defined by the zero level of proton, it is qualified as a proton donor) (Dickson, 1981; Wolf-Gladrow et al., 2007). Evidently, the zero level of protons does not mean that there is zero proton in the solution (Wolf-Gladrow et al., 2007). For the carbonate acid-base system at pH=4.5, carbonic acid is the dominant species and therefore defines the zero level of protons, from which the number of released or uptaken protons to reach that level are deduced. With this respect, HCO_3^- can accept one proton (Equation 1.9 is pushed to the left) and CO_3^{2-} can accept two protons (explaining the coefficient 2 in Equations 1.12 and 1.13). With carbonic acid as the zero level of protons, bicarbonate and carbonate ions are therefore defined as proton acceptors (Wolf-Gladrow et al., 2007)[4]. Whilst the car-

[4]For example, if the reference level was chosen at pH=8, bicarbonate ion would defines the zero level of protons, and with this respect, would quality carbonic acid as a proton donor and carbonate ion as a proton acceptor, which would lead to the coefficients +1 and -1 in front of the corresponding concentrations in the TA equation, respectively.

bonate species account for more than 95% of TA (Williams and Follows, 2011), the other acid-base systems present in seawater play an important role in fully defining TA (Zeebe and Wolf-Gladrow, 2001). For each acid-base system, the classification of proton acceptors and proton donors is also determined on their corresponding level of zero protons at pH=4.5. The excess of proton acceptors over proton donor for all acid-base systems is therefore included in the TA definition (Equation 1.12) (Dickson, 1981).

$$TA = [\mathrm{HCO_3^-}] + 2[\mathrm{CO_3^{2-}}] + [\mathrm{B(OH^-)_4}] + [\mathrm{OH^-}] + [\mathrm{HPO_4^{2-}}] + 2\,[\mathrm{PO_4^{2-}}] + [\mathrm{H_3SiO_4^-}] + [\mathrm{NO_3}] + [\mathrm{HS^-}] - [\mathrm{H^+}] - [\mathrm{HSO_4^-}] - [\mathrm{HF}] - [\mathrm{H_3PO_4}] + \cdots \quad (1.12)$$

The conservative expression of TA (Equation 1.13) is however useful to understand how certain physical and biological (Section 1.2.2) processes can impact TA (Williams and Follows, 2011). For instance, a change in the dilution of the major ions' concentration present in seawater, through precipitation/evaporation events, would induce a decrease/increase in surface TA, respectively. As such, surface TA is proportional to salinity, with relatively high values in the subtropical regions or in the Mediterranean Sea (areas of high evaporation) and relatively low values in the subpolar regions (areas of high precipitations). Disparities between the ocean basins, for example the relatively lower TA in the Pacific Ocean than in the Atlantic, is also explained by initial difference in the basins' salinities (the Pacific Ocean being less salty than the Atlantic).

$$[\mathrm{Na^+}] + 2\,[\mathrm{Mg^{2+}}] + 2\,[\mathrm{Ca^{2+}}] + [\mathrm{K^+}] - [\mathrm{Cl^-}] - 2\,[\mathrm{SO_4^{2-}}] - [\mathrm{Br^-}] - [\mathrm{F^-}] - [\mathrm{NO_3^-}] = TA = [\mathrm{HCO_3^-}] + 2[\mathrm{CO_3^{2-}}] + [\mathrm{B(OH^-)_4}] + [\mathrm{OH^-}] - [\mathrm{H^+}] + \cdots \quad (1.13)$$

The Revelle factor

While all the above properties have so far been introduced independently, they interact through their response to rising atmospheric CO_2 concentration (among other things) via the Revelle factor. The Revelle factor describes the sensitivity of $pCO_{2\text{-ocean}}$ to change in DIC (Equation 1.14) (Sarmiento and Gruber, 2006). In other words, the Revelle factor characterises the buffering capacity of seawater and therefore the ability of seawater to take up additional CO_2: a high Revelle factor indicates a low buffer capacity (i.e. lower ability to

take up additional CO_2), while a low Revelle factor indicates a high buffer capacity (i.e. higher ability to take up additional CO_2). Warm tropical and subtropical waters typically have a low Revelle factor, while the cold high latitudes have a high Revelle factor (Figure 1.6) (Sabine et al., 2004). As such, under rising atmospheric CO_2 concentrations, high latitudes regions (e.g. the subpolar North Atlantic) would reach a saturation in the air-sea CO_2 flux more rapidly than subtropical regions (assuming no major biology or circulation change) (Halloran et al., 2015).

The buffer capacity in seawater is specifically explained by the neutralisation of the protons (H^+), which are released through the dissolution of CO_2 in seawater, by the bicarbonate and carbonate ions (amongst other ions) (Williams and Follows, 2011). The abundance of carbonate species in seawater is the primary explanation for the strength of the buffer effect in the oceans (Zeebe and Wolf-Gladrow, 2001). Since fresh water systems (such as lakes) have low concentrations of carbonate and bicarbonate ions, they are weakly buffered. For instance, if 1 µmol·kg^{-1} of HCl solution is added to a fresh water body initially at pH = 7, T = 15 °C and DIC = 2,000 µmol·kg^{-1}, a drop of 1 pH unit is observed, while it would lead to a reduction of 0.003 in seawater (at S = 35 g·kg^{-1}) (Zeebe and Wolf-Gladrow, 2001). As such, the amplitude of the Revelle factor is proportional to the ratio between DIC and TA, or more specifically is a function of these two variables (right hand side of Equation 1.14) (Sarmiento and Gruber, 2006)).

$$\Gamma = \frac{\delta p\mathrm{CO}_{2-\mathrm{ocean}}}{p\mathrm{CO}_{2-\mathrm{ocean}}} \frac{[DIC]}{\delta[DIC]} \approx \frac{3 \cdot TA \cdot DIC - 2 \cdot DIC^2}{(2 \cdot DIC - TA)(TA - DIC)} \tag{1.14}$$

1.2.2 Biological processes

While the surface ocean is of great interest as it is in direct contact with the atmosphere and hence is where exchange of anthropogenic CO_2 occurs, the deep marine chemistry ultimately plays a crucial role in closing the carbon cycle, especially through key biological processes. Indeed, the biological pump allows the sequestration of carbon to the ocean depth and by doing so, impacts atmospheric CO_2 concentrations. Carbon is specifically fixed by biology in the surface waters through two different reactions: photosynthesis and calcification. The cycle of the carbon through these two processes into the abyss is referred as the soft-tissue pump and the carbonate pump. Here, the soft-tissue and the carbonate pumps are presented,

and their impacts on the properties previously developed (i.e. $fCO_{2\text{-ocean}}$, DIC and TA) are described.

The soft-tissue pump

The soft-tissue pump broadly corresponds to the cycling of organic matter that is formed in the surface waters into inorganic compounds at depth. Organic matter is formed by photosynthesis, a net effect generated by phytoplankton organisms (Equation 1.15) (Williams and Follows, 2011).

$$6CO_2 + 6H_2O + \text{nutrients} \xrightarrow{\text{photons}} C_6H_{12}O_6 + 6O_2 \tag{1.15}$$

By consuming CO_2 and nutrients such as nitrate (NO_3^-), photosynthesis leads to a decrease in DIC and an increase in TA (Equation 1.13 shows an increase in TA when the concentration of nitrate decreases) proportional to the ratio of carbon to nitrogen of -117:16 (Figure 1.7) (Sarmiento and Gruber, 2006). The decrease in DIC in the euphotic layer (i.e. the uppermost sunlight layer) results in a decrease in surface $fCO_{2\text{-ocean}}$ (Figure 1.7), which therefore tends to drive a local CO_2 uptake (Williams and Follows, 2011). The carbon that is fixed into organic matter follows two possible routes:

1. It remains in the surface layers and returns to its inorganic form through bacteria activities or respiration of zooplankton (De La Rocha, 2007). Respiration is the opposite reaction of photosynthesis, which therefore corresponds to the oxidation of organic matter and leads to an increase in DIC and a decrease in TA (Figure 1.7) (Williams and Follows, 2011).

2. It is exported to the deep ocean through advection or sinking of particles (Sarmiento and Gruber, 2006). Nevertheless, only a small fraction (about 5-10%) of the sinking particles (the relatively large particles, that mostly result from optimal coagulation conditions or the packaging of material into faecal pellets) eventually reach the seafloor and therefore contribute to the burial of carbon into sedimentation (Buesseler, 1998; De La Rocha, 2007). The rest of the sinking particles are decomposed by microbial activities and/or zooplankton grazing throughout the water column (De La Rocha, 2007). The remineralisation of organic carbon into inorganic carbon therefore contributes to

the increase of DIC with depth. Those enriched DIC intermediate/deep waters eventually return to the surface through specific oceanic circulation patterns (e.g. upwelling).

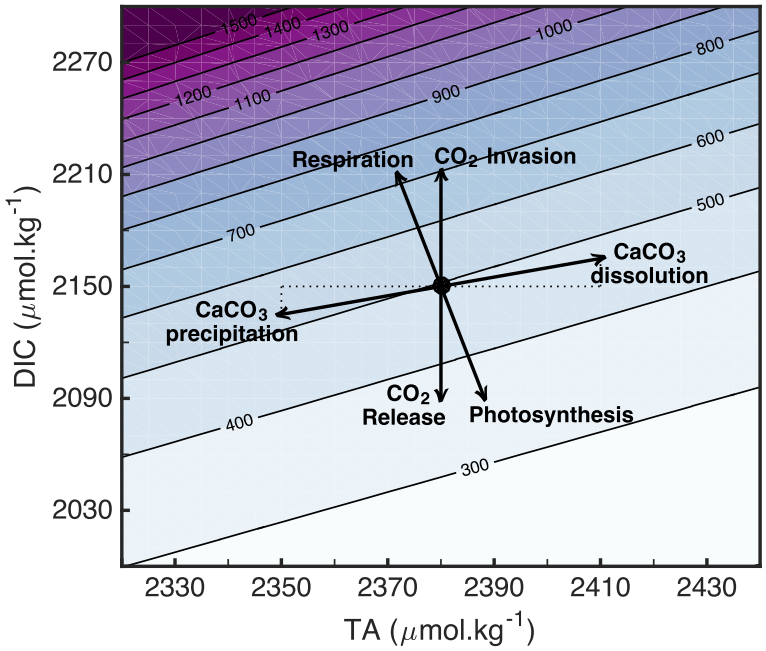

Figure 1.7: Impact of biological processes and CO_2 invasion on marine chemistry. DIC, TA and $fCO_{2\text{-ocean}}$, which is given in µatm by the isolines, were calculated using the CO2SYS Matlab toolbox, for a salinity of 35 and a temperature of 20°C (Van Heuven et al., 2011; Sarmiento and Gruber, 2006). The ratio 1:2 in DIC:TA involved with $CaCO_3$ formation/dissolution is highlighted by the dotted lines.

The efficiency of the soft-tissue pump depends on the efficiency of primary production (i.e. the amount of organic matter that is produced by photosynthesis) (Buesseler, 1998). One of the criteria for active primary production is the availability of nutrients in the euphotic layer. Nutrient-enriched deep waters are regularly brought-up to the euphotic zone through intense vertical motion, such as coastal upwelling, or confined to the near-surface through the strengthening of the thermocline in spring (Geider et al., 1998). Localised increase in the near surface nutrients supply, assuming no corresponding increase in DIC supply, therefore leads to a decrease in DIC and an increase of the air-sea CO_2 flux (Williams and Follows, 2011). Under global warming conditions, while the increase of sea surface temperature would generally be expected to increase phytoplankton growth and therefore stimulate primary production, the induced increase in stratification would reduce vertical mixing and hence decrease the supply of nutrients in the euphotic layer (Denman et al., 2007). In nutrient limiting regions (i.e. low latitudes), such impacts could potentially lead toward a decrease

in the soft-tissue pump's efficiency (Kwiatkowski et al., 2017). However in light limiting regions (i.e. high latitudes), a narrowing of the initially nutrient-enriched mixed layer could concentrate phytoplankton into the euphotic layer, which would enhance primary production and thus increase the soft tissue pump's efficiency.

The carbonate pump

The carbonate pump corresponds to the cycling of calcium carbonate ($CaCO_3$) from the surface to the deep ocean. Calcium carbonate is a mineral formed through the calcification reaction (Equation 1.16) generated by coccolithophores or other calcareous organisms. Each mole of $CaCO_3$ precipitated leads to the decrease of one mole of DIC and two moles of TA (due to the stoichiometric number 2 in front of the carbonate ion concentration in Equation 1.13; Figure 1.7) (Zeebe and Wolf-Gladrow, 2001). As such, calcification increases $fCO_{2\text{-ocean}}$, which can either reduce the air-sea CO_2 flux or lead to a localised CO_2 outgassing to the atmosphere.

$$Ca^{2+} + {CO_3}^{2-} \longleftrightarrow CaCO_3 \tag{1.16}$$

Once precipitated in the surface layers, calcareous shells sink into intermediate/deep waters. As in the soft-tissue pump, only a small fraction of calcareous shells reach the seafloor and is buried into sedimentation. Indeed, most of the sinking shells are progressively dissolved due to the increase of their solubility with depth (i.e. solubility increases with lower temperature and higher pressure), contributing to a vertical increase in DIC and TA (Figure 1.7) (Williams and Follows, 2011). The depth at which calcium carbonate changes from supersaturation (i.e. favourable to precipitation) to undersaturation (i.e. favourable to dissolution) state is referred as the saturation horizon (Williams and Follows, 2011). Under rising atmospheric CO_2 concentrations, the saturation horizon is expected to become shallower, due to a decrease in pH that reduces ${CO_3}^{2-}$ concentrations and therefore promotes the dissolution of calcium carbonate. For example, model projections show that in the North Atlantic the saturation horizon for aragonite (i.e. a mineral consisted of $CaCO_3$) will shift from 2850 m to 150 m depth by 2100 under the business-as-usual scenario (Ciais et al., 2013). Among other repercussions, a shallower saturation horizon would impact DIC profiles, leading to higher subsurface DIC concentrations and therefore potentially intensified CO_2 outgassing following vertical

mixing events. The reduction in surface calcification from ocean acidification could also decrease the ability of the oceans to store CO_2 at depth (Feely et al., 2004).

In summary, the two biological pumps (i.e. soft-tissue and carbonate) provide similar changes in DIC throughout the water column, although with different depth profiles, while TA varies in opposite directions (i.e. there is a decrease in TA due to the remineralisation of organic matter at depth and an increase in TA due to the dissolution of calcium carbonate with depth).

The various interactions between biology and chemical/physical properties highlight the great complexity of the marine carbon cycle, and as such, the challenges that the modelling community can face when representing this cycle in biogeochemical models (e.g. sinking rate, horizon saturation, nutrient availability). Nevertheless, even if models successfully reproduce key aspects of the marine carbon biogeochemistry, they may fail to represent those interactions in a temporally varying climate system and regionally heterogenous oceanic circulation. Indeed, the combination of specific oceanic features with various modes of atmospheric variability in certain regions can lead to further impacts on the oceanic CO_2 uptake, which is the case for the North Atlantic ocean; a key region for climate change mitigation.

1.3 The CO_2 uptake in the North Atlantic

The North Atlantic is a region of strong contemporary CO_2 uptake (Sabine et al., 2004; Mikaloff Fletcher et al., 2007; Khatiwala et al., 2013; example of the anthropogenic CO_2 column inventory in Figure 1.4). The contemporary air-sea CO_2 flux is described by two components: (1) the natural CO_2 component, which existed at pre-industrial times and is mostly driven by circulation and biology natural changes, and (2) the anthropogenic CO_2 component, which is driven by the atmospheric CO_2 response to anthropogenic activities and the oceanic feedbacks due to climate change (Mikaloff Fletcher et al., 2007). The intense North Atlantic contemporary CO_2 sink is closely related to the specific oceanic circulation in that basin, as well as the strong biological pump occurring in the subpolar region (Schuster and Watson, 2007). To understand the challenges in quantifying the spatially and temporally-varying North Atlantic CO_2 uptake, this section introduces the main surface and deep circu-

lation features contributing to the CO_2 uptake (Section 1.3.1) and describes how the modes of variability of that basin influence that sink (Section 1.3.2). Within this context, the recent past change in the North Atlantic CO_2 uptake is discussed (Section 1.3.3). The sometimes contrasting results between studies will highlight the difficulty of providing robust CO_2 uptake estimates and will therefore put into context the objectives of this thesis developed at the end of this chapter (Section 1.6). The North Atlantic boundaries for the rest of the thesis are defined from 10°N to 70°N and from -75°E to 5°E (except for independent studies which may use a different region boundary).

1.3.1 The North Atlantic circulation

The North Atlantic surface circulation is mainly wind-driven and is described by an anticyclonic (i.e. clockwise in the Northern Hemisphere) subtropical gyre and a cyclonic (i.e. anticlockwise) subpolar gyre (Figure 1.8). The subtropical gyre is a warm and saline region, while the subpolar is cold and relatively fresh from sea-ice melt and high precipitation. The zone of transition between the two gyres is highly variable, which is mainly due to the strong baroclinic instabilities generated by the Gulf Stream, a narrow (less than 120 km wide) and strong (up to 140 Sv, where 1 Sv $= 10^6$ $m^3\,s^{-1}$) western boundary current (Talley et al., 2011). The Gulf Stream feeds the North Atlantic Current (NAC), a warm and saline current that flows eastward, forming the northern and southern boundaries of the subtropical and subpolar gyres, respectively (Figure 1.8; Talley et al., 2011). Due to a strong bottom topography steering in the subpolar gyre, the NAC separates into two main branches: a northward branch, which is known as the Norwegian Current or North Atlantic Drift, and an eastward branch which then meets the western side of the subpolar gyre (Talley et al., 2011). The cyclonic circulation of the subpolar gyre induces a divergent Ekman transport and consequently the upwelling of deep cold waters, which under specific conditions contributes to the formation of deep waters.

In the open ocean, deep water convection is initiated by the isopycnal outcropping at the surface (for instance through Ekman divergence), which allows deep waters to be in contact with the cold atmosphere, especially in winter (Marshall and Schott, 1999). The vigorous winter oceanic heat loss creates a strong buoyancy forcing, which increases vertical density instability and hence stimulates vertical mixing (Kuhlbrodt et al., 2007). Once the localised

strong forcing ceases, lateral exchange dominates vertical heat transfer, allowing the advection of the newly formed dense water on the horizontal plane (Kuhlbrodt et al., 2007). In the North Atlantic, the Labrador Sea is the main region of open ocean deep water formation, where Labrador Sea Water sinks up to 2000m depth and recirculates in the western side of the Labrador basin (Marshall and Schott, 1999; Kuhlbrodt et al., 2007). The Greenland and Irminger Seas are also key regions of open ocean deep convection, where ice formation and localised atmospheric conditions play an important role in the convection preconditioning phase (Kuhlbrodt et al., 2007; Weaver et al., 2007; Pickart et al., 2003). The specific features of the subpolar North Atlantic sets this region as one of the few convective regions in the world (Talley et al., 2011).

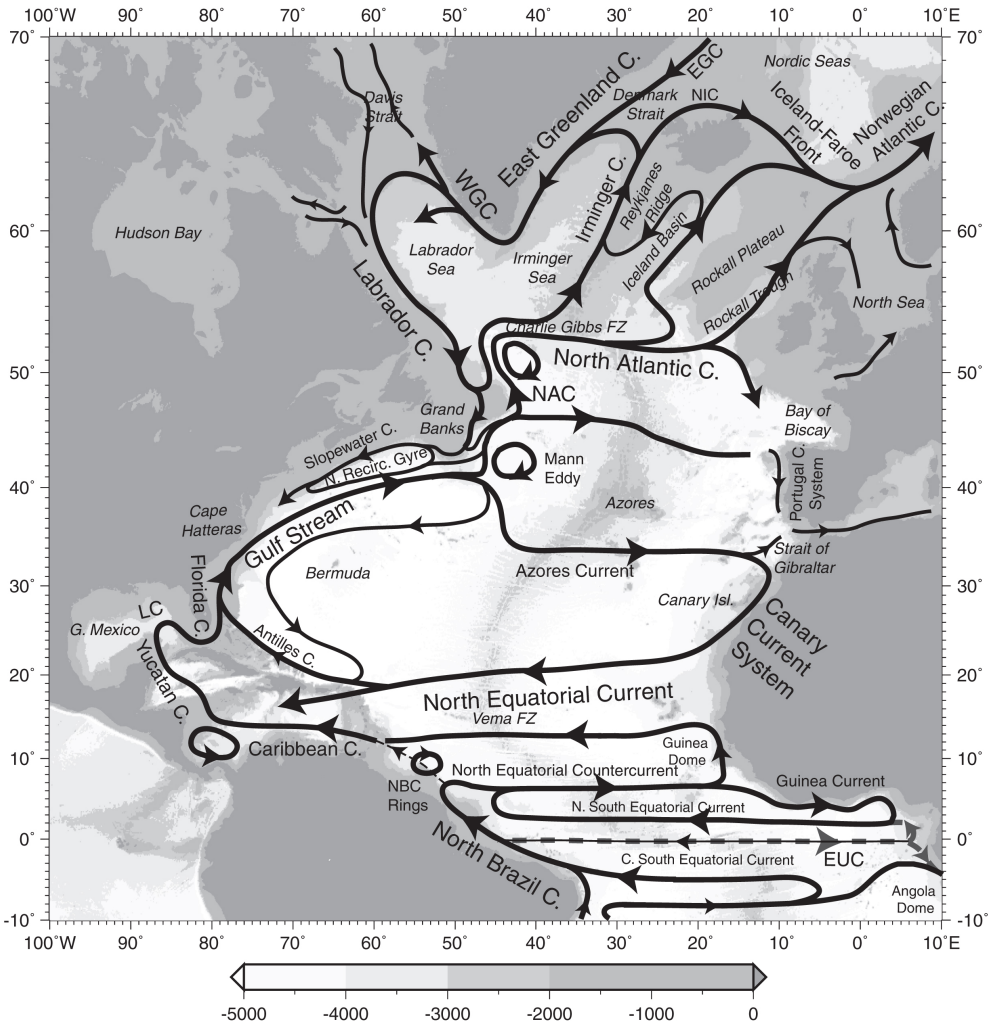

Figure 1.8: North Atlantic Circulation. Representation of the currents in the North Atlantic. Source: Talley et al. (2011).

The surface circulation and deep water formation of the North Atlantic explain well the intensity of the CO_2 uptake occurring in that basin (Figure 1.5). First, surface waters experience on their northward journey abrupt thermal and biogeochemical gradients (Halloran et al., 2015). As surface waters' temperature decreases from the subtropical to the subpolar gyre, the solubility of CO_2 increases and the speciation between the carbonate species changes, enhancing the air-sea CO_2 flux (Section 1.2.1) (Halloran et al., 2015). In the subpolar region, intense biological activity removes carbon from the surface ocean, increasing the air-sea CO_2 gradient and triggering local CO_2 uptake, particularly on seasonal timescales. For instance, observations in the northeast Atlantic recorded an intensification of the air-sea CO_2 flux in summer, which was explained by a stimulation of photosynthesis at that time of year (Kortzinger et al., 2008). Indeed, as the mixed layer shallows with increasing temperature, nutrients are brought up to the euphotic layer, which therefore enhances photosynthesis (Kortzinger et al., 2008). Potentially before reaching equilibrium with the atmospheric CO_2, the carbon-enriched subpolar surface waters are then removed into the intermediate/deep ocean through deep convection events or the formation of mode waters, which allow the North Atlantic to maintain surface waters with a strong air-sea CO_2 gradient, and consequently air-sea CO_2 flux (Takahashi et al., 2009; Halloran et al., 2015).

1.3.2 Internal climate variability

While the large-scale mechanisms controlling the North Atlantic carbon sink are well-understood, processes linked to internal climate variability add further complications. The interactions between internal climate variability and the CO_2 uptake are not yet fully understood due two main reasons: (1) the relatively short length of the observational CO_2 record prevents the capture of all modes of variability complete internal cycles and therefore limits full investigation of the mechanisms of variability (e.g. McKinley et al., 2011); (2) natural climate variability is superimposed on top of the anthropogenically-forced climate, which are difficult to differentiate using observational data alone. By contrast, model analyses provide the advantage to be able to fully quantify the model's internal climate variability from control simulations, but the internal variability that the model simulates highly depends on the model's configuration and will not necessarily match that occurring in the real-world. Here, two features of internal climate variability in the North Atlantic are

presented and their interactions with the CO_2 sink, mostly from model studies, are discussed.

Variability within the Atlantic Meridional Overturning Circulation

While the Atlantic Meridional Overturning Circulation (AMOC) describes a general circulation feature, the variability in the strength of that circulation influences the cycle of CO_2 between the deep and surface waters. The AMOC is made of four branches: the northward surface flow, the deep water convection at high latitudes, the southward deep currents that spread dense water masses into the ocean interior and the upwelling branch that brings deep waters back to the surface (e.g. through the wind-driven upwelling in the Southern Ocean, mixing processes). The AMOC is specifically characterised by the strength of the overall northward transport throughout the water column.Variations in the AMOC strength impact the meridional transport of heat from the low to high latitudes (Cunningham et al., 2013), the surface carbon chemistry and consequently the biology, which feed back on the formation of deep waters and the storage at depth of anthropogenic carbon-enriched surface waters. A model-based study suggested that decadal change in the recent global CO_2 uptake was primarily due to variability in the strength of the upper oceanic circulation (DeVries et al., 2017). For instance, the study proposes that the observed weakening of the overturning circulation in the 2000s led to a reduction in (1) the upwelled DIC-enriched deep waters, which suppressed the outgassing of natural CO_2, and (2) the formation of intermediate/deep water, which limits the storage of anthropogenic CO_2 at depth (DeVries et al., 2017). As such, an increased of the natural CO_2 sink (0.6 PgC·yr^{-1}) and a decrease in the anthropogenic CO_2 (0.1 PgC·yr^{-1}) was identified in the 2000s; leading to an overall increased in the global contemporary ocean CO_2 sink (DeVries et al., 2017). Other model studies (e.g.: Gregory et al., 2005; Weaver et al., 2007) identified that the weakening of the AMOC might be intensified in the future under rising atmospheric CO_2 concentrations, which could lead to further relative decrease in the storage at depth of anthropogenic CO_2.

The North Atlantic Oscillation

The North Atlantic Oscillation (NAO) describes a change (relative to the mean) in sea level pressure between the subtropical high and subpolar low (Talley et al., 2011). The NAO index is commonly defined by the pressure difference between Portugal and Iceland (Talley

et al., 2011). Variations in the phase of the NAO index mainly occur from inter-annual to decadal timescales and impact atmospheric circulation, as well as physical and biogeochemical oceanic properties. For instance during a positive phase of the NAO, the increased pressure difference (i.e. stronger subtropical high and subpolar low relative to the mean state) leads to an intensification and northward shift of the westerly winds (Hurrell et al., 2003). This atmospheric shift particularly perturbs the eastern side of subpolar gyre, through the northward extension of the subtropical gyre and the acceleration of the NAC. As the NAC accelerates, the supply of low-carbon concentrations and warm waters on the eastern side of the subpolar gyre increases, which intensifies the CO_2 uptake and weakens the deep water formation in the Greenland Sea, respectively (Gruber et al., 2009; Talley et al., 2011). Oceanic perturbations also occur on the western side of the subpolar gyre. The intensification of the Labrador Current increases the southward transport of high-carbon concentration and cold waters, which weakens the CO_2 uptake near the Canadian coast and strengthens the Labrador Sea Water production, respectively (Gruber et al., 2009; Talley et al., 2011). For instance, Pérez et al. (2013) suggested that during the high-NAO phase of the early 1990s, the storage of anthropogenic CO_2 was about three times greater than during the low-NAO phase of 2002-2006. This intensification of the carbon storage could be explained by the vigorous uptake in the eastern part of the subpolar gyre and the exceptional convection events that occurred in the Labrador and Irminger seas during the early 1990s (Pérez et al., 2013).

Due to the highly variable nature of the North Atlantic, analyses that study trends in the surface CO_2 uptake over relatively short time-periods (from inter-annual to decadal timescales) may capture internal variability rather than responses due to rising atmospheric CO_2 concentrations (McKinley et al., 2011). It is therefore essential to sustain carbon observations in order to build-up long-term data record, which will allow us to deepen our understanding of the the carbon cycle response to oceanic and atmospheric changes.

1.3.3 Recent change in the North Atlantic CO_2 uptake

The current carbon-observing system is made of *in-situ* measurements (unlike satellite-based data) and is described by two types of datasets: (1) time-series records made

at a "fixed" location, which provide useful information on the change in ocean carbon content but are ultimately constrained geographically; and (2) measurements made at various locations and at various times from research cruises and Voluntary Observing Ships (VOS), which provide a relatively wide spatial coverage but would lead to further challenges when quantifying the temporal change of the carbon system. Despite international efforts to provide an improved CO_2 observational coverage (both spatially and temporally), the remaining gaps in the coverage are tackled differently within the community, leading occasionally to contradicted trends in the recent North Atlantic CO_2 sink between studies (e.g. Schuster and Watson, 2007; Lefèvre et al., 2004; Schuster et al., 2013), which by using different methods, and sometimes time periods and basin boundaries lead to a difficult inter-comparison assessment.

Repeated measurements of the seawater carbonate chemistry at a "fixed" location allows one to study the change in the carbonate system at frequencies that are not yet achieved with large-scale measurements. For instance at the Bermuda Atlantic Time-series Study (BATS) station, measurements (DIC, TA, temperature and salinity) are made at the same location 14-15 times a year (Bates et al., 2012). The BATS station combined with a no-longer active station (i.e. the hydrostation S, which was operating from 1983 to 1988, while the BATS station has been operating since 1988) provides the longest time series of the marine carbonate system (Bates et al., 2012). Using these data, a significant increase of about 20% in surface $pCO_{2\text{-ocean}}$ was detected during 1983-2011, which corresponds to a comparable rate of increase in atmospheric pCO_2 (Bates et al., 2012). As such, in the subtropical Atlantic (or at least at the BATS station) the CO_2 sink has not significantly changed over the period 1983-2011 (Bates et al., 2012). However this result neglects the influence of the winds on the air-sea CO_2 flux, which were known to have intensified in the 2000s potentially due to the shift of the winter NAO index.

Large-scale surface carbon measurements are mostly obtained from underway instruments deployed on VOS. As such, the surface $fCO_{2\text{-ocean}}$ data distribution is skewed to the path of commercial shipping routes and times when sea conditions are favourable, leaving certain regions/months where no carbon data are recorded. In the North Atlantic, the UK-Caribbean shipping route is particularly active and the data collected through that line was the focus of Schuster and Watson (2007). Using data from two VOS, one covering the period 1994-1995

and the other 2002-2005, Schuster and Watson (2007) observed an increase in the annual $pCO_{2\text{-ocean}}$ between the two periods (a mean increase of 4.4 µatm·yr^{-1} if assuming a linear trend), leading to an overall decrease in the CO_2 ocean sink of 0.24 PgC·yr^{-1} in the region between 20°N and 65°N. The variation in the NAO between the mid-1990s to the early 2000, through the change in the surface temperature and ventilation, was proposed to be the main cause for the reduction in the CO_2 sink (Schuster and Watson, 2007); a result that was also suggested by Pérez et al. (2013) and consistent with additional analysis in Schuster et al. (2009). However when using longer time periods, other studies (e.g. Takahashi et al., 2009) show that the North Atlantic $pCO_{2\text{-ocean}}$ increase at a rate that is indistinguishable from the rate of increase in $pCO_{2\text{-atmosphere}}$, and as in Bates et al. (2012), found that there was no significant change in the oceanic CO_2 sink over a three decade coverage.

While measurements in the North Atlantic were made in the 1970s (as used in Takahashi et al., 2009 for their trend calculations), the sparsity of these data and their potential lack of precision/accuracy open questions regarding their reliability and their meaningful contribution to the overall signal, especially if they were sampling a phase of internal variability. As mentioned above, some of the first trend studies established their conclusions on the changing North Atlantic CO_2 sink based on a $\sim$10-15 year long period (e.g. Schuster and Watson, 2007; Schuster et al., 2009), which most-likely captures decadal-scale variability rather than long-term emergent trends responding to the increasing atmospheric CO_2 concentrations (McKinley et al., 2011). Indeed, McKinley et al. (2011) illustrated through an observation-based analysis across three biogeographic regions of the North Atlantic that it takes about 25 years for the anthropogenic driving force to dominate the ocean carbon uptake. However in the subpolar biome, they highlighted that further multi-decadal climate variability (e.g. the Atlantic Multidecadal Variation, whose proposed period is about 60 years) may still be influencing the $pCO_{2\text{-ocean}}$ trends across their period of study (i.e. 1981-2009). While McKinley et al. (2011) based their trend calculations on $pCO_{2\text{-ocean}}$ observations from the 1980s, when observations were much more limited than the last two decades, they evaluated the impact of the irregular data coverage on the deduced trends using a model. Specifically, they sampled a regional physical-biogeochemical model at the same locations and times at which measurements were made, calculated the trend using the sampled model field and compared it with the original model trend (i.e. the trend

calculated using the model field before sampling). As such, they could evaluate in each biome if the data coverage would overestimate or underestimate the trend that would have been returned if the oceans were perfectly sampled, and their model performed realistically. While such a novel approach makes use of the strength of the observational system and the modelling world, the robustness of the study relies on the behaviour of one specific model and on the fact that this model provides identical biogeographic regions and modes of variability as in the real ocean, which may well not be the case.

While the carbon system is sensitive to different timescales and study intervals, the spatial coverage in $pCO_{2\text{-ocean}}$ observations also plays an important role when assessing the basin-wide North Atlantic CO_2 sink. To overcome the irregular distribution of surface $pCO_{2\text{-ocean}}$ observations, various gap-filling methods[5] have been developed (e.g. multiple linear regression in Watson et al., 2009; neural network in Landschützer et al., 2013; autocorrelation in Jones et al., 2015). For example, Schuster et al. (2013) extended their above work on the North Atlantic ship tracks to the mapping of surface carbon observations on the Atlantic and Arctic regions from 1990 to 2007 (alongside with various observation-based and model products). Their method relied on a MLR approach, which fitted an optimal relationship between the surface $pCO_{2\text{-ocean}}$ and other physical and biogeochemical fields that are available at every location and step in time (e.g. satellite-derived products, climatological products). Using this relationship on the complete physical and biogeochemical fields, they were able to reconstruct basin-wide surface $pCO_{2\text{-ocean}}$. However since one of the predictor variables was not available in the Arctic and during the entire study period (i.e. 1990-2007), their analysis had to be constrained to 1998-2007 and to the Atlantic only; a detail that highlights the cost of depending on other observational products and how as a result, the trend based on the basin-wide $pCO_{2\text{-ocean}}$ was omitted in the study (as it did not match the period from the other products). They nevertheless provided trends of the spatially integrated air-sea CO_2 flux for another interpolated observational product (Takahashi et al., 2009), which showed an overall increase in the oceanic CO_2 sink in the subpolar and subtropical North Atlantic regions for the period 1995 to 2009; a signal that is in agreement with other modelling products (the intensity of the sink however differs across the methodologies).

[5]Gap-filling methods will be discussed in depth in Chapter 4.

In summary, while the North Atlantic is a highly sampled region, temporal and spatial gaps in the carbon-data coverage lead to challenges in the assessment of the change in ocean CO_2 uptake. Decadal studies identified a decline in the North Atlantic subpolar CO_2 sink during the mid-1990s to the early 2000s potentially due to internal variability (e.g. Schuster and Watson, 2007), while long-term studies (of about 30-year long; Takahashi et al., 2009, Bates et al., 2012) however seem to show a non significant change in the North Atlantic CO_2 uptake. The diversity in results from observation-based studies clearly highlights some of the limitations when using observational-data only. Models on the other hand allow the investigation of drivers, long-term processes and responses that would not be feasible in the real-world, such as exploring the evolution of North Atlantic CO_2 into the future under climate change conditions, particularly through the CMIP5 framework.

1.4 The CMIP5 framework

The CMIP5 framework is an internationally coordinated set of simulations generated by multiple state-of-the-art Earth System Models (ESM)s, aiming to improve the current understanding of past and future climates (Taylor et al., 2012). More than 20 modelling groups using more than 50 models took part within the CMIP5 framework, offering a wide range of responses to similar future socio-economic scenarios; responses that were particularly used for the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) (Taylor et al., 2012). The results generated by the CMIP5 simulations have allowed (among other things) to the community to (1) deepen the understanding of the various interactions within the Earth system (e.g. Jones et al., 2013), (2) explore the reasons behind models' diverse responses to similar forcing, and (3) constrain future projections based on real-world observations (e.g. Kwiatkowski et al., 2017). The CMIP5 models, which will be extensively used throughout this thesis, are therefore introduced here with the presentation of key definitions (Section 1.4.1) and are used to initiate the discussion around the future change in the North Atlantic CO_2 uptake (Section 1.4.2).

The specific use of the CMIP5 models throughout this thesis, instead of for example high resolution models, is justified by the fact that the CMIP5 framework provides a large model

diversity and freely accessible data, which will allow us in Chapter 4 (the core methodological chapter) to statistically investigate in depth the uncertainty associated with a mapping interpolation technique used in the real-world, and therefore constrain for the first time the recent change in the observed North Atlantic CO_2 uptake. Nevertheless, while the model diversity in the CMIP5 framework contributes to the robustness of the interpolation uncertainty assessment (c.f. Chapter 4), depicting and understanding a common behaviour across the CMIP5 models remains challenging. As such, Chapter 5 will introduce a set of simulations from a forced ocean-only model, which will allow us to explore some of the reasons behind a significant bias in the CMIP5 models in capturing the observed North Atlantic CO_2 sink. The forced ocean-only model will therefore contribute to key side analysis to understanding the CMIP5 models and will be presented in due time (c.f. Chapter 5, Section 5.3.1). In this introduction chapter, the focus is therefore made on the CMIP5 models, which will be used in all chapters of this thesis. A description of the specific CMIP5 models used in this thesis, as well as the associated data processing, are developed in Appendix A.1.

1.4.1 Definitions and set-up

Earth System Model

An ESM is a fluid-dynamical representation of the atmosphere and ocean, in which non-linear interactions between the ocean, atmosphere, cryosphere and biosphere are described by a set of physical, chemical and biological equations, including the carbon cycle and feedbacks to the climate system (Figure 1.9) (Collins et al., 2011; Tjiputra et al., 2013). Unlike climate models, ESMs include (among other additional components) a dynamical representation of the land and oceanic carbon cycle, interacting and responding to anthropogenic forcing and climate change feedbacks. For instance, one of the CMIP5 models, the Met-Office HadGEM2-ES, incorporates various interacting processes such as the emissions of dimethyl sulfide by plankton into the atmosphere (Halloran et al., 2010), which form particles that interact with cloud formation and hence the Earth's radiative balance (Figure 1.9). ESMs are therefore useful tools allowing testing of our theoretical understanding of the Earth system and generating future climate projections, but are ultimately constrained by our current understanding of the processes involved and expense of computing resources (Collins et al., 2011).

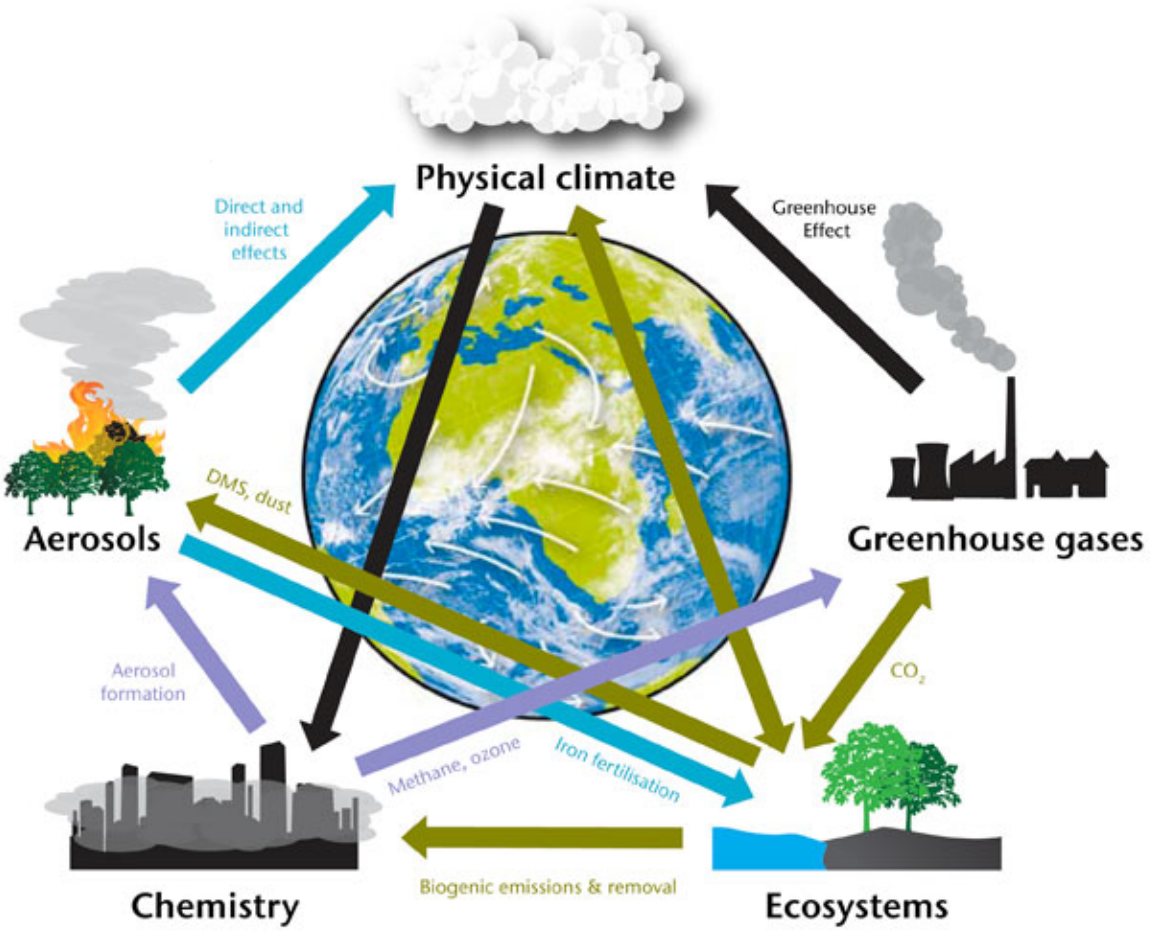

Figure 1.9: Schematic representation of an Earth System Model. Specific representation of the HadGEM2-ES model. Source: http://www.metoffice.gov.uk/research/news/2012/cmip5

Biogeochemical models

The biogeochemical component of an ESM is typically included to allow simulation of the carbon cycle (Bopp et al., 2013). The biogeochemical component commonly follows the four-compartment model type Nutrient Phytoplankton Zooplankton Detritus (NPZD), to which various levels of complexity are added (Bopp et al., 2013; Table A.2).

A NPZD model describes via a set of parameters and relationships the broad features of the biological pump (e.g. phytoplankton growth/mortality, zooplankton grazing/excretion, detrital sinking and remineralisation) (Scott et al., 2011). The sensitivity to the parametrisation (i.e. the value that is given to a parameter) on the modelled air-sea CO_2 flux was studied by Scott et al. (2011). In their study, Scott et al. (2011) emulated a wide range of values for all parameters within the NPZD model HadOCC (Palmer and Totterdell, 2001) and found that (at three sites in the North Atlantic) the air-sea CO_2 flux was mostly influenced by the pa-

rameters controlling phytoplankton growth, sinking rate of detritus and carbonate formation. For instance, through the efficiency of phytoplankton at fixing DIC, surface DIC content is modified, which therefore impacts surface $pCO_{2\text{-ocean}}$ and air-sea CO_2 flux (Scott et al., 2011). Such example therefore highlights the importance of the representation and parametrisation of the biological processes on the simulated oceanic CO_2 uptake.

To embrace some of the complexity of ocean biogeochemistry, Phytoplankton Functional Types (PFT) models have been developed. A PFT model uses for each type of phytoplankton a different set of parameters aiming to describe their specific behaviour (Scott et al., 2011). Most of the biogeochemical models used in the CMIP5 ESMs have adopted a PFT scheme, representing from one to three types of phytoplankton (e.g. diatoms, flagellates) and/or zooplankton (e.g. microzooplankton, mesozooplankton), as well as a different number of limiting nutrients (Table A.2; Bopp et al., 2013). While the addition of key biogeochemical features could potentially lead to more realistic projections, and importantly allows for more non-linear behaviour to occur through the interaction of multiple species, the limits to our understanding of the ecology and lack of data are likely to lead to inaccurate parametrisation and increased uncertainty (Anderson, 2005; Scott et al., 2011), which could feed back the source of uncertainties when predicting the future CO_2 oceanic sink. Also, by representing specific species rather than by characterising a theoretical average phytoplankton as in a NPZD model, a PFT model ultimately implies that those are the only important species in the carbonate system. One of the limitations in current biogeochemical models (NPZD and PFT) is the lack of interactions between the ecosystem behaviour and the environmental change (Scott et al., 2011), where for example calcification will not be affected by ocean acidification (Bopp et al., 2013). Whether the pathways forward is to add more complexity in biogeochemical models or improve the parametrisation of simply represented processes (both of which present their advantages and drawbacks) therefore remains a debate within the modelling community.

Experiments

To describe how the Earth system will evolve under anthropogenically-driven perturbations, ESMs are forced with a range of prospective socio-economic, air pollutants/particulates and emission of greenhouse gases scenarios called Representative Concentration Pathways (RCP)

(Van Vuuren et al., 2011), that also include natural forcing emissions (e.g. volcanic aerosol emissions) (Jones et al., 2013). To facilitate the comparison of the models' response(s) to climate change, a common set of four different concentration-driven RCPs, each of them corresponding to a possible climate impact and mitigation scenario, were prescribed to the CMIP5 models (Figure 1.10; Van Vuuren et al., 2011). By specifically prescribing globally averaged greenhouse gases concentrations (as opposed to emissions), carbon-cycle feedbacks do not interfere with the atmospheric CO_2 concentrations and as such, the focus is on comparing the impact of four different climates on the various components of the Earth system.

The four concentration-driven scenarios are labelled RCP8.5, RCP6, RCP4.5 and RCP2.6,

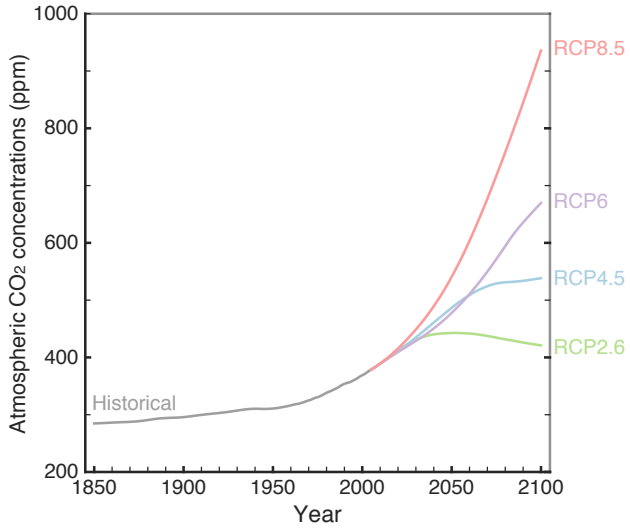

Figure 1.10: Atmospheric CO_2 concentrations driving scenarios. Time series of the annual globally averaged atmospheric CO_2 concentrations that is used to force the historical and the respective RCP scenarios simulations.

whose numbering corresponds to the estimated maximum radiative forcing by 2100 (relative to pre-industrial conditions) (Figure 1.10; Van Vuuren et al., 2011). For example, the radiative forcing in the RCP8.5 scenario, also known as the "business-as-usual" scenario, reaches a level of about 8.5 $W{\cdot}m^{-2}$ at the end of the 21st century relative to pre-industrial levels (Riahi et al., 2007; Taylor et al., 2012). On the other hand, the radiative forcing in the RCP2.6 scenario, also referred as the "peak-and-decline" scenario, reaches a maximum near the middle of the 21st century before decreasing due to mitigation actions to a level of 2.6 $W{\cdot}m^{-2}$ relative to pre-industrial values (Van Vuuren et al., 2007).

During the historical period (in CMIP5 experiments considered to be from 1850 to 2006), the experiments were forced with atmospheric concentrations reproducing the observed natural

and anthropogenic changes (Figure 1.10; Taylor et al., 2012). Throughout this thesis, the historical and RCP8.5 scenario will be used, as well as the RCP2.6 scenario on few occasions (e.g. Figure 1.11, Chapter 6).

1.4.2 The future of the North Atlantic CO_2 uptake

In order to inform policy makers on the maximum CO_2 emissions permitted to remain a certain threshold of warming by 2100, it is essential to evaluate the future responses of the carbon reservoirs, including the North Atlantic CO_2 sink, to climate change conditions and feedbacks. Here, the CMIP5 models are used to briefly discuss the potential fate of the future North Atlantic CO_2 sink.

The future North Atlantic CO_2 sink is studied through two atmospheric CO_2 concentration scenarios: the "business-as-usual" scenario (i.e. RCP 8.5) and the "lower emission" scenario (i.e. RCP 2.6) which describes the impact of mitigation and adaption actions (red and green lines in Figure 1.11, respectively; Taylor et al., 2012; Riahi et al., 2007). While the ESMs show relatively good agreement on the 21st century global ocean CO_2 uptake, their behaviour at the North Atlantic scale differ (Figure 1.11) (Halloran et al., 2015). The modelled North Atlantic CO_2 sink also clearly displays pentadal to multidecadal variability, illustrating the challenge of quantifying trends over a relatively short interval (discussed in Section 1.3.3). The trajectory of the North Atlantic (and global ocean) air-sea CO_2 flux clearly follows the atmospheric CO_2 scenarios pathways: as CO_2 increases/decreases in the atmosphere, more/less CO_2 is absorbed by the surface ocean (Figures 1.11 and 1.10). Nevertheless, while the business-as-usual scenario follows an on-going increase in atmospheric CO_2 concentrations (Figure 1.10), the CO_2 oceanic sink seems to reach a plateau around 2075 for the global ocean and to be followed by a decline for the North Atlantic (Figure 1.11), which could be explained by a change in the carbonate speciation in seawater in response to the CO_2 invasion into seawater and/or by the fact that the gradient of atmospheric CO_2 becomes linear toward the end of the century (Zeebe and Wolf-Gladrow, 2001; Halloran et al., 2015; Figure 1.10). Through a box model study, Halloran et al. (2015) analysed the drivers of a "peak and decline" response of the subpolar North Atlantic CO_2 sink in a business-as-usual scenario. As in Völker (2002), they identified that the difference in the Revelle factor between the subtropical and subpolar (i.e. low Revelle factor at the low

latitudes and high Revelle factor at the high latitudes) mostly explain the "peak and decline" feature occurring under increasing atmospheric CO_2 concentrations. Indeed, the high buffer capacity in subtropical waters induces a strong CO_2 uptake within the 21st century and when northward transport of those carbon-enriched waters begins to satisfy the carbon uptake capacity of the low buffered subpolar region, the air-sea CO_2 uptake at the high latitudes starts to decline.

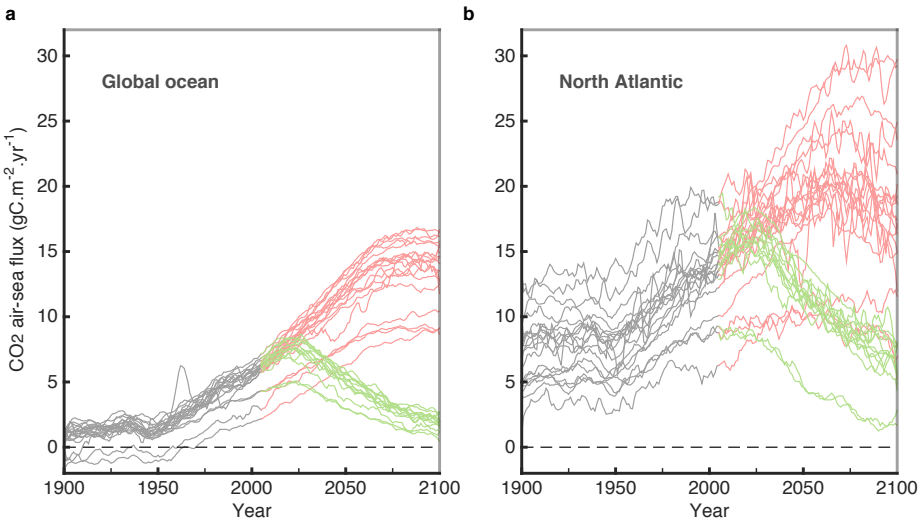

Figure 1.11: Simulated oceanic CO_2 uptake. Annually varying **a**, Global, **b**, North Atlantic air-sea CO_2 flux (positive meaning a flux into the ocean) using model outputs for the historical experiment (grey) and for the business-as-usual (red) and mitigation/adaptation (green) scenarios (see Section 1.4.1 for further information on these simulations). Averages were area-weighted and filtered with a 5-year running mean. Each individual line corresponds to an individual model.

To assess where the real-world ocean stands within the "peak and decline" timeline and therefore assess whether the subpolar North Atlantic has already started to reach saturation, it is crucial to robustly quantify the recent change in the CO_2 oceanic sink (which will be tackled in Chapter 4), as well as maintaining *in-situ* observations. However, due to the change in the driving mechanisms between short and long-term timescales, understanding the mechanisms of the recent change in the North Atlantic CO_2 uptake may not directly inform us how this sink is likely to evolve in the future (Halloran et al., 2015). It is therefore critical that we also improve the representation of the carbon cycle within ESMs.

1.5 Challenges of model evaluation

While future projections are essential to understand the responses of the Earth system to climate change conditions, their reliability can never be guaranteed. The common way to gauge how good those projections are is to evaluate the simulated physical and biogeochemical fields against our current knowledge of the real-world; a crucial step referred as model evaluation. Since model evaluation will be a major focus on Chapters 2, 3 and 5, the challenges in evaluating the results of the models' simulations against observation-based products are here discussed into four main points:

- The unknown truth. Models are usually compared with observations while asking "How well does the model represent the truth?" (Stow et al., 2008). One main issue with this approach is that the truth remains unknown, as it cannot be measured (Stow et al., 2008). As such, the predictive error (i.e. the difference between the model prediction and the truth) and the observational error (i.e. the difference between the observation and the truth) are unknown (Figure 1.12). The true state of the system is therefore assumed to lie within the observational uncertainty (Stow et al., 2008), and as will be discussed in Chapters 2, 3 and 4, quantifying the observational uncertainty is also very challenging. The unknowns behind both models and observations can therefore lead to difficulties when interpreting comparative results.

- Limited observations. Model evaluations are constrained to regions and times at which measurements were conducted. Certain fields (e.g. oceanic carbon-related variables) present relatively limited amount of observations, which could potentially mislead the interpretation of the model skill (i.e. the model's performance in reproducing the real-world). For instance, a model can present a good skill based on localised observations, without guaranteeing that it will capture the globally and temporally varying system (e.g. seasonal, decadal variabilities).

- The choice of the evaluation method(s). Different level of investigation behind model evaluation studies (from common basic comparisons to extensive statistical analysis) can lead to different interpretation of the model's performance (Stow et al., 2008). For example, the Taylor diagram is a commonly-used and efficient communication tool, which displays in a compact manner three statistical quantities (i.e. correlation,

standard deviation and *unbiased* Root Mean Square Error (RMSE)) (Taylor et al., 2001; Jolliff et al., 2009). However, a Taylor diagram can display a model close to ideal statistical values while this model can actually have a significant bias compared to the observations (Jollif et al., 2009; Friedrichs et al., 2009). A model evaluation analysis that would solely rely on Taylor diagrams might therefore miss important model biases. To reveal all the different aspects of model skill, the use of complementary metrics is therefore highly recommended (Stow et al., 2008).

- Scenarios uncertainty. Even if a model presents a high skill against current observations, it does not necessarily mean that its future projections are more plausible than the projections from other models. Indeed, model evaluation cannot assess the skill of future climate feedbacks and variability simulated within those models, which are not only constrained by the model set-up, but by the uncertainties behind the RCP scenarios (Booth et al., 2012; Halloran et al., 2015).

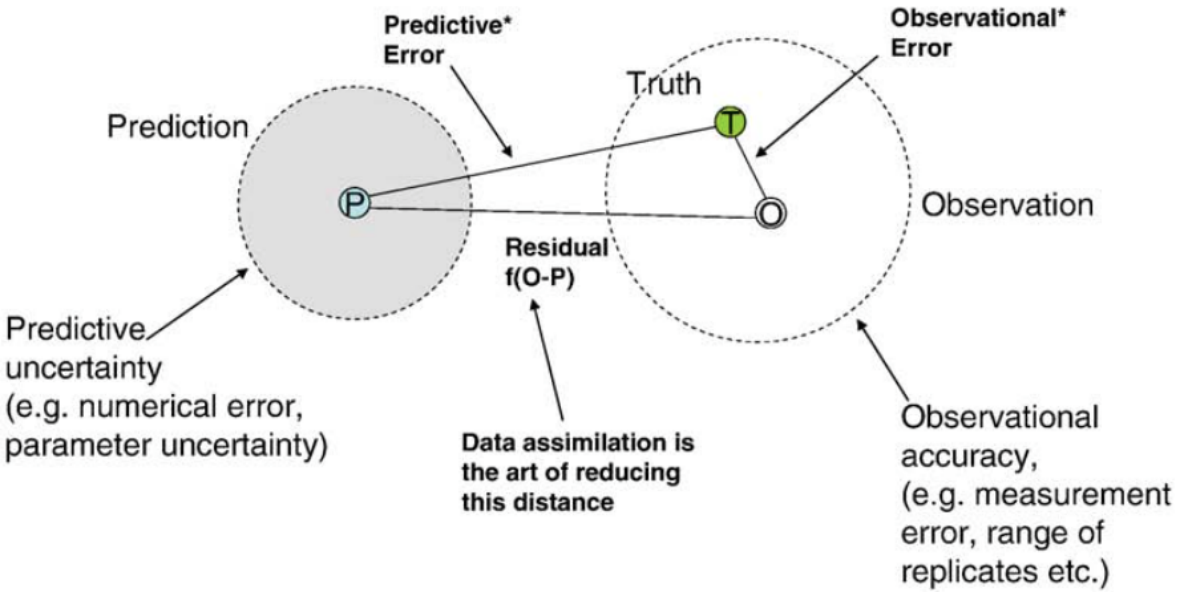

Figure 1.12: Context behind model evaluation. Relationships between observations (O), model prediction (P) and the unknown truth (T). The star refers to unknowns, as the true state cannot be measured. Dotted circles represent the predictive (grey) and observational (white) uncertainties. A model has skill when those two uncertainties overlap. Source: Stow et al., (2008).

1.6 Thesis plan

In summary, through an intense surface anthropogenic CO_2 uptake and storage at depth, the North Atlantic Ocean plays a key role in mitigating against the on-going rising atmospheric CO_2 concentrations. Quantifying recent change in the North Atlantic CO_2 sink is essential to (1) understanding its contribution to the global oceanic uptake and thus to the global carbon budget, (2) determining whether carbon cycle models are performing well, and (3) assessing the degree to which the sea surface saturation in CO_2 has begun, which is key to understanding how much CO_2 we can rely on this sink taking up in the future. Nevertheless, in a highly variable system and with relatively limited oceanic carbon observations, getting a complete and robust picture of change in the North Atlantic CO_2 sink and explaining the drivers of this change remain challenging. To tackle these challenges, the scientific community has been (1) using models, which provide the advantage of being perfectly defined at each grid point and time-step, but ultimately only represent a virtual version of the real-world; and (2) developing gap-filling methods on carbon observations, which despite being based on real-world information add an additional unknown source of uncertainty. The main aim of this Ph.D. is therefore to deepen our understanding of the recent changing North Atlantic CO_2 sink by combining the strengths of both observational and modelled products. To reach this aim, it is essential to first fully grasp the complexity behind both observational and modelled products. As such, Chapters 2 and 3 focus on evaluating the biogeochemistry of the latest generation of ESMs (i.e. the CMIP5 models) against observational-based products (specifically DIC and TA for Chapter 2 and $fCO_{2\text{-ocean}}$ for Chapter 3), but also on highlighting some of the sources of uncertainties behind observational products that are often treated as the "truth". With this context, Chapter 4 presents a method based on models and the current observational coverage that determines, in a robust manner, the uncertainty associated with the interpolation of observational $fCO_{2\text{-ocean}}$. Chapter 5 then puts into perspective the deduced well-constrained trend in the North Atlantic CO_2 uptake with trends from the CMIP5 models, but mostly investigates with a set of simulations the potential reasons behind the identified systematic bias in the CMIP5 models. Finally, the conclusion chapter (Chapter 6) summarises the main results raised in all chapters and discusses the potential future implications of the current CO_2 sink underestimation identified in the CMIP5 models.

Chapter 2

Evaluating Dissolved Inorganic Carbon and Total Alkalinity in Earth System Models

This chapter aims to evaluate DIC and TA in the CMIP5 models against the Global Ocean Data Analysis Project (GLODAP) version 2 observational product, highlighting important potential biases in the models that will be examined in Chapter 5.

GLODAPv2 is a widely-used product that summarises discrete oceanic *in-situ* measurements of a selection of physical and biogeochemical variables into interpolated climatologies (Key et al., 2015; Lauvset et al., 2016; Olsen et al., 2016). The basin-wide nature of this product has been particularly appreciated by the modelling community for model evaluation purposes as it allows for a straightforward comparisons with model outputs (e.g. Bopp et al., 2013; Ilyina et al., 2013). However, the comparison of model data with an interpolated observational product that include various sources of uncertainties lead to further challenges when robustly assessing the models' skills. To provide a meaningful model skill assessment on the DIC and TA fields simulated by the CMIP5 models, it is therefore essential to understand and investigate the sources of uncertainty behind GLODAPv2. As such, this chapter first introduces and discusses the various challenges when summarising 42 years of *in-situ* DIC and TA measurements into interpolated climatologies (Section 2.1) and within that context, assesses the North Atlantic DIC and TA climatologies of 15 available CMIP5 models (Section 2.2).

By highlighting the importance of providing robust basin-wide uncertainties on interpolated

observational product, this chapter will introduce the main motivation behind Chapter 4, which will determine the North Atlantic basin-wide uncertainties on the time-varying surface $fCO_{2\text{-ocean}}$.

2.1 Investigating the GLODAPv2 observational error

GLODAPv2 is the result of extended international collaborations and many years of building knowledge. To put GLODAPv2 into perspective and to fully appreciate the observational uncertainties behind the mapped DIC and TA climatologies, the various stages leading that product are explored, with a particular focus on the challenges associated with producing high-quality measurements (Section 2.1.1). The impact of the improvements made in GLODAPv2 compared to its previous version (e.g. increase of measurements, technical improvements) on model evaluations is then discussed, particularly through the example of a model evaluation study (Bopp et al., 2013) (Section 2.1.2). This example particularly emphasises the importance of not treating observational-based products as the true state. Finally through a model-based study, the impact of the seasonal sampling bias on climatologies, which is one of the main acknowledged remaining biases in GLODAPv2, is explored (Section 2.1.3).

2.1.1 DIC and TA measurements: distributions and uncertainties

Temporal and spatial distributions

Over the past decades, joint international efforts have led to increased *in-situ* oceanographic measurements throughout the basins and depths, providing an extremely valuable resource for documenting the changing state of the oceans (Talley et al., 2016). The first international oceanographic collaboration was achieved in the 1970s with the Geochemical Ocean Sections (GEOSECS) programme (Craig and Gordon, 1972). While this collaborative initiative has set the baseline of our current knowledge of marine biogeochemistry (e.g. Stuiver et al., 1983), the lack of robust accuracy standards at the time led to occasionally inaccurate DIC and TA measurements (Key et al., 2004; Olsen et al., 2016). In the 1990s, the World Ocean Circulation Experiment (WOCE) global survey significantly contributed to the extension of repeated hydrographic sections and to the use of a variety of devices (e.g. floats, moorings) (WOCE, 2002). To provide a useful and high-quality baseline for future

climate studies, the WOCE community developed throughout the course of the programme specific standards for precise and accurate measurements. For carbon measurements, the Standard Operation Procedure (SOP)s and Certified Reference Material (CRM)s (Dickson et al., 2007) were specifically developed around 1995, providing to this day the highest degree of precision and accuracy standards (Key et al., 2004). More recently, the Climate and Ocean: Variability, Predictability and Change (CLIVAR) / Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP) survey has also defined hydrographic sections to be repeated on a decadal frequency, in order to quantify the interior decadal variability of the seawater CO_2 system, among other oceanographic variables (Talley et al., 2016).

The limited temporal and spatial coverages of the various research programmes, as well as their different measurement standards between the periods prior and after the SOP and CRM era, has led to relatively heterogenous biogeochemical *in-situ* data. To provide a useful DIC and TA dataset for model evaluation studies, as well as for understanding the current state of the global ocean carbon cycle, the main challenge for the GLODAP community was therefore to uniformly merge these measurements. Below, the temporal and spatial distributions of the DIC and TA measurements incorporated in GLODAPv2 are described.

From 1972 to 2013, a total of 689,279 DIC and TA measurements were globally integrated in GLODAPv2, of which $\sim$22% were taken from the top 100 m of the water column and $\sim$56% measured in the Northern Hemisphere (Figure 2.1). The North Atlantic is a particularly highly sampled region, while other regions (the Indian and Southern Oceans) present relatively sparse data. Over the 42 years of data collection, more DIC than TA measurements were made on average (Figure 2.1c). Peaks in the number of measurements are observed in the mid 1990s and mid 2000s (Figure 2.1c), indicating the important contribution of the WOCE and CLIVAR/GO-SHIP global surveys, respectively. Interestingly, the number of measurements dropped slightly prior to 2000 while there is no corresponding drop in the number of voyages (Figure 2.1b-c). The lack of consistent correlation between the numbers of voyages and measurements (Figure 2.1b,c) also highlights the "trade off" between improving the spatial gaps across the basins (usually by increasing the number of voyages) and improving the horizontal and vertical resolution in a transect (by increasing the number

of measurements). However, the mismatch observed prior 2000 could have also been due to more specific reasons, such as poor weather conditions leading to fewer stations or poor quality data which were then flagged out[1]. Moreover, a decrease in the number of voyages is observed since 2009, corresponding to a drop in the number of measurements. While the number of voyages is not expected to drop due to the sustained research programmes (although it might due to the difficulties in funding), this decrease may be due to delay in making the collected data publicly available and hence for the GLODAP community to incorporate those in their product.

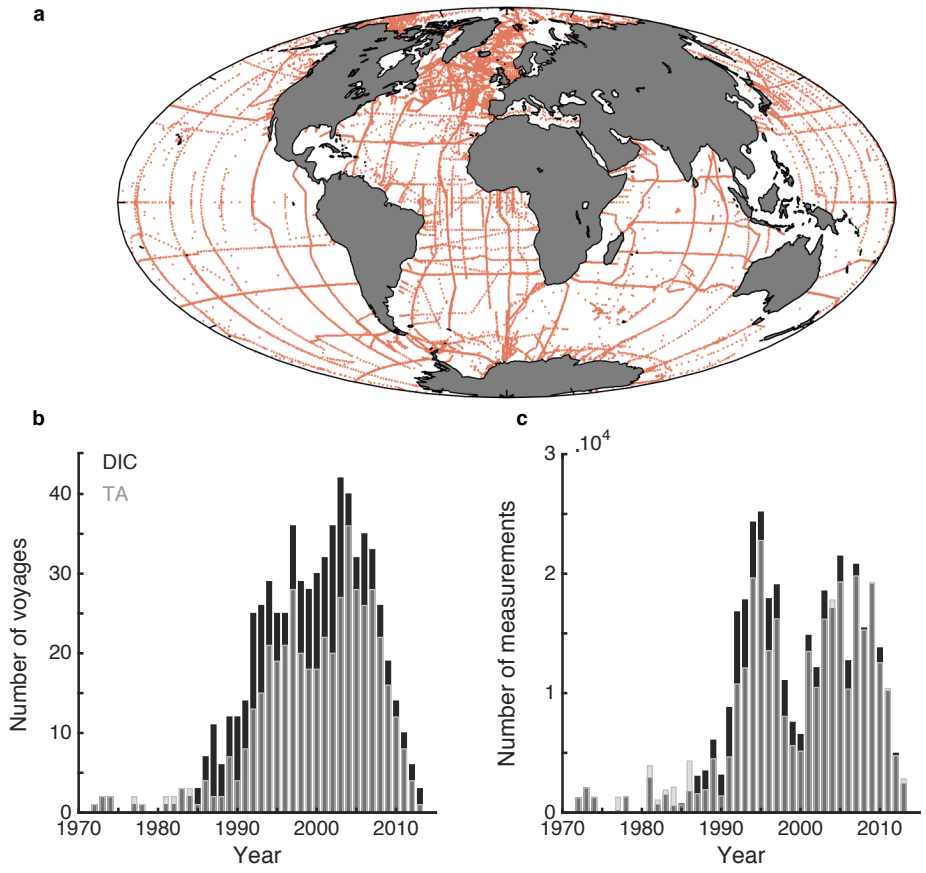

Figure 2.1: Spatial and temporal distribution of carbonate measurements used in GLODAPv2. **a**, global spatial distribution of both DIC and TA measurements used in GLODAPv2, including data collected throughout the water column. **b**, annual number of research voyages that measured DIC (dark grey) and TA (light grey) . **c**, annual number of all DIC (dark grey) and TA (light grey) measurements.

[1]For full investigations behind the specific reasons of that drop in the measurements, please refer to each corresponding voyage reports.

Providing high-quality data

Another crucial point to take into account when dealing with observations is the determination of the precision and accuracy of a measurement. How reproducible is a measure within the same experimental design? How close is the measurement to the (unknown) true value? Although the uncertainty in the input data is not the major source of uncertainty in GLODAPv2 (the interpolation error is; Lauvset et al., 2016), I believe that to appreciate in greater depth the level of complexity behind such product, one needs to understand the various steps to produce high-quality DIC and TA measurements[2]. Here I describe those various steps:

- **The sampling strategy.** Unlike some properties that are measured continuously through the water column (e.g. temperature, salinity), DIC and TA measurements are made from discrete seawater samples collected from a rosette, which usually holds either 12 or 22 Niskin bottles. As such, the vertical sampling is limited at each station to the number of Niskins available. The depths at which the Niskin bottles are closed, are commonly selected to capture the main features identified from the continuous profiles (e.g. change of water masses, peak of chlorophyll), with usually increased resolution toward the surface. In practice, it is unfortunately common to have a few faulty Niskins, which either repetitively fail to close or leak. To adjust to this, the on-board team would usually close another Niskin to the same depth as the unpredictable bottle, which guarantees this specific depth to be sampled but ultimately decreases the vertical sampling resolution. The first challenge to provide a representative "snapshot" of the carbon system is therefore to find an appropriate horizontal/vertical sampling resolution, which is constrained by the number of available Niskins, the duration of the voyage and the often limited time allocated for samples analysis.

- **The first Quality Control.** The first level of Quality Control (QC) aims at identifying outliers due to technical or instrumental issues and at discerning the "Acceptable" to the "Questionable" data, following the WOCE flag protocols. If samples are anal-

[2]The measurements challenges are specifically mentioned here due to the laboratory experience I gained during two research voyages. The developed quality control work of one of the voyage is developed in Appendix B.

ysed on-board, the initial first QC can be carried out during the analysis (e.g. use of salinity to highlight potential faulty samples), allowing any technical or instrumental issues to be addressed immediately. However in practice, the first QC is not always straightforward, particularly when unexplained technical issues occur (c.f. Appendix B.2 for examples). The limited on-board time prevents exhaustive laboratory investigations and as such, the carbon scientists are faced with essential questions: If those unexplained issues cannot be fixed for the time-being, how to get around them? How to treat the measurements made when those issues occurred? The first QC is therefore relatively conditioned by the scientific team, their experience and sometimes their "intuition" in identifying questionable data or suspicious instrumental behaviour.

- **The precision analysis.** The precision of the DIC and TA measurements is determined through the comparison of duplicate results (SOP 2, 3; Dickson et al., 2007). Due to time constrains and laboratory costs, duplicate analysis can only achieved on few samples per station (Dickson et al., 2007; Appendix B.3). The on-board duplicate comparison also allows the identification of potential issues with the machines which may be immediately resolved. However, a duplicate can return two significantly different results and one can then wonder: Is this difference due to poor precision or to a specific experimental issue? Which one of the two values is likely to be the closest to the unknown truth? The precision can also vary quite substantially between machines with identical experimental design, which contradicts the identical precision across the machines provided by the manufacturer. Determining the precision of a measurement is therefore a challenging task, that is device-specific, and sometimes may even be seawater-specific (Appendix B.3).

- **The calibration.** For highly accurate calibration purposes, Certified Reference Materials (CRMs) have been created in 1995 and used ever since (A. Dickson, Scripps Institution of Oceanography, USA). Note that prior to 1995, calibration techniques were used but did not followed the current standards. CRMs from the same batch contain identical seawater that was sterilised through filtration, ultra-violet radiation and poison addition and that was accurately measured through exhaustive methods (Dickson et al., 2007). The DIC and TA values provided by the batch reference, which are (questionably?) treated as the truth, are compared to the values returned by the devices. This

comparison allows the quantification of the instrumental offset and hence, provides appropriate calibration (Appendix B.4). Also due to time and cost constrains, CRMs cannot be analysed between each samples. As such, the instrumental offset is typically calibrated to the nearest-in-time analysed CRMs, which may not always be the most appropriate approach (particularly when there is a significant drift in the instrumental offset).

- **The second level of Quality Control.** The second level QC identifies, through a crossover analysis, potential systematic biases in the collected data with reference to historical measurements (Lauvset and Tanhua, 2015). Specifically, the crossover analysis compares the deep water (i.e. usually below 1,500 m depth) properties from the studied voyage with previous voyages in a 2° radius. Such comparison assumes that (1) within the observational-record timescale, the deep water properties remain invariant (Lauvset and Tanhua, 2015), and (2) the previous voyages are bias-corrected; both which are questionable assumptions. For instance, a crossover analysis showed that a recent voyage presents a significantly different bias depending on whether pre-CRM or post-CRM voyages were used (Appendix B.5). This result could either be due to remaining biases in the pre-CRM voyages or that the CO_2 system in the deep waters has significantly evolved since. With maintained *in-situ* measurements into the future, one could therefore question the assumption of invariant properties at depth, particularly in regions of deep water formations or deep western boundary currents.

  The second QC also depends on the locations of the crossovers. For instance, if the studied voyage is only crossed on one side of the section, the identification of a systematic bias throughout the entire voyage could be even more challenging. One could therefore imagine that in the future, more evenly-spread voyages across the basins could improve crossover analyses, and that the second QC of that particular voyage could be re-evaluated. Differences in the crossover analyses could then be due to increased number of data in the reference database or also because the reference voyages are closer in time to the studied voyage than in an earlier version of GLODAP. Crossover analyses are therefore relative to the information gathered by the reference voyages, leading to a systematic unknown when assessing the precise bias of a voyage.

Overall, providing high-quality DIC and TA measurements implies a series a technical and evaluation steps, that are often influenced by the scientists' judgement and constrained by assumptions that could be questionable. The integration of measurements taken from various different devices, experimental procedures, standards and ultimately made by scientists with potentially different experience, can therefore be a task of great complexity. As such, the GLODAP community unified the precision of the entire data to a single value (Key et al., 2004). To adjust the potential systematic voyage biases on various oceanographic properties, extensive second QC was also performed within GLODAPv2. Once those steps achieved, all in-situ data are binned into a regular grid and then interpolated, providing basin-wide estimates of the DIC and TA climatologies (Lauvset et al., 2016).

2.1.2 Perspectives on the interpolation error

While the synthetic and interpolated features in GLODAP are valuable for model evaluations, it also feeds its main source of uncertainty (Lauvset et al., 2016). For instance, the interpolation technique used in GLODAPv2 seems to underestimate the real errors by 25% (Beckers et al., 2014; Lauvset et al., 2016). To appreciate the impact of the increased number of measurements and improvements in the interpolation methods on the resulting mapped climatologies, I present the two main versions of GLODAPs (versions 1 and 2, Table 2.1) and discuss how these different versions can impact model skills[3], using the study of Bopp et al., (2013) as an example.

GLODAP version	Voyages	Horizontal resolution	Depth levels	Climatology period	Variables	Mapping Method	References
1	116	$1° \cdot 1°$	33	1972-1999 except DIC, CFC $<$ 1.2 km : WOCE only	TA, DIC potential alkalinity anthropogenic CO_2 $\Delta^{14}C$ (bombed & natural) CFC-11, CFC-12 pCFC-11, pCFC-12	Optimal Interpolation	Key et al., 2004
2	724	$1° \cdot 1°$	33	1972-2013, for all variables except: pH, Ω_C, Ω_A, $<\sim$1 km: 1986-1999 & 2000-2013 DIC normalised in 2002	DIC, TA, pH Ω_C, Ω_A nitrate, phosphate, silicate salinity, theta	DIVA	Key et al., 2015 Olsen et al., 2016 Lauvset et al., 2016

Table 2.1: Main characteristics of GLODAPv1 and GLODAPv2. Extended comparison between the two GLODAP versions is described in Lauvset et al. (2016) and Olsen et al. (2016).

[3]While the two GLODAP versions represent two different mean states (as version 1.1 is dominated by the WOCE surveys in the 1990s and version 2 by the CLIVAR/GO-SHIP survey in the 2000s), I assume for comparative purposes that this difference is relatively minor as they both share the common goal of representing the global and climatological state of the ocean biogeochemistry.

GLODAP version 1

GLODAP version 1 (Table 2.1) was released in 2004 and was at the time the first globally and vertically integrated ocean carbon product available (Key et al., 2004). This product contributed to key scientific findings, such as the Sabine et al. (2004) study that identified the North Atlantic as the region of highest anthropogenic carbon concentration integrated through the water column on the planet. GLODAPv1 also contributed to the first global biogeochemical model evaluations (e.g. Zahariev et al., 2008; Bopp et al., 2013). For instance, a model inter-comparison study (Bopp et al., 2013) used the surface DIC and TA climatologies from GLODAPv1 to determine the surface pH and evaluated the pH skill of the models (Figure 2.2). Note that the surface pH (unspecified pH scale) calculated in Bopp et al. (2013) is hereinafter referred as pH_{v1} (Figure 2.2b). The models were found to underestimate the spatial variability of the observation-based pH_{v1} (particularly in the North Atlantic, North Pacific and Southern Ocean), and hence return an overall poor correlation with the global observation-based pH_{v1} value (Figure 2.2) (Bopp et al., 2013).

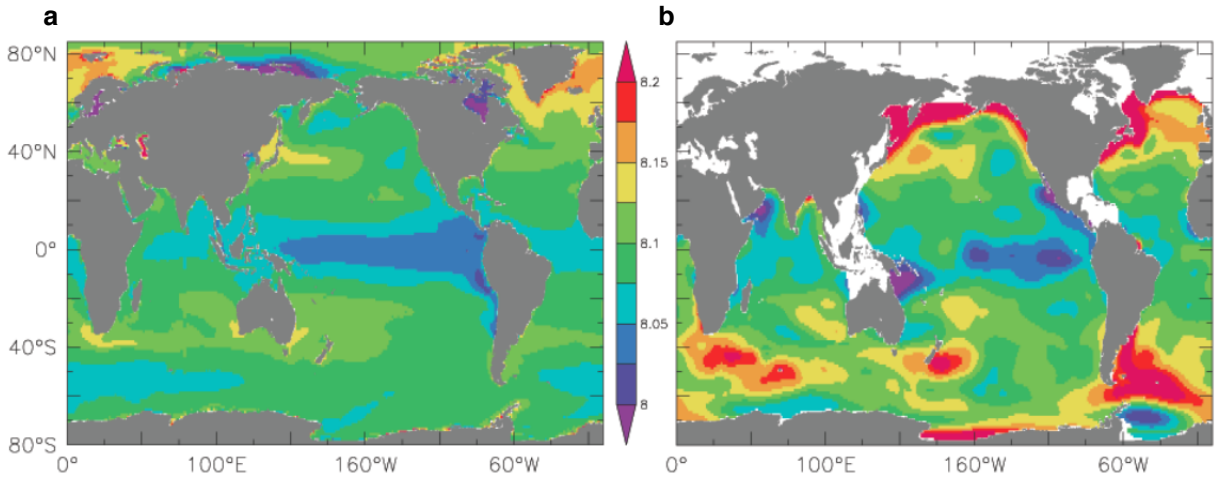

Figure 2.2: Model-observation comparative study of surface pH. a, model-mean surface pH climatology calculated from model outputs between 1990 and 1999. **b**, pH climatology calculated from DIC and TA climatologies from GLODAPv1 (representative of the 1990s). Source: Bopp et al. (2013).

The model evaluation made in Bopp et al. (2013) was based on treating the deduced observational pH as the "true" pH state and as such, one could wonder if the poor pH skill is either due to the models generally failing to represent the real-world marine carbon chemistry (i.e large residual in Figure 1.12) or due to (unknown) large observational error

(i.e. large observational error in Figure 1.12). Indeed, the skill results could be altered by the fact that GLODAPv1 (1) is a climatology, which by definition represents a mean state and not the true state, (2) is an interpolated product with an interpolation error which is potentially underestimated (Key et al., 2004) (note that in Bopp et al. (2013) the interpolation error was not considered), (3) partially uses DIC and TA measurements that did not follow the current high-quality standards (although all data within GLODAP have been quality controlled), and more importantly (4) does not provide mapped pH climatology based on binned pH values. Due to the non-linearities of the carbonate system, the production of pH climatology is indeed sensitive to the ordering of the calculation steps: a pH climatology calculated from the DIC and TA interpolated products (Figure 2.3a) will differ from a pH climatology that has interpolated binned pH values, themselves calculated from binned DIC and TA pairs (Figure 2.3b). Consequently, is the high spatial variability in pH calculated from the GLODAPv1 climatologies (Figure 2.2b) due to calculations ordering artefact or represent the real state of the pH surface ocean? This question is investigated using the GLODAPv2 product.

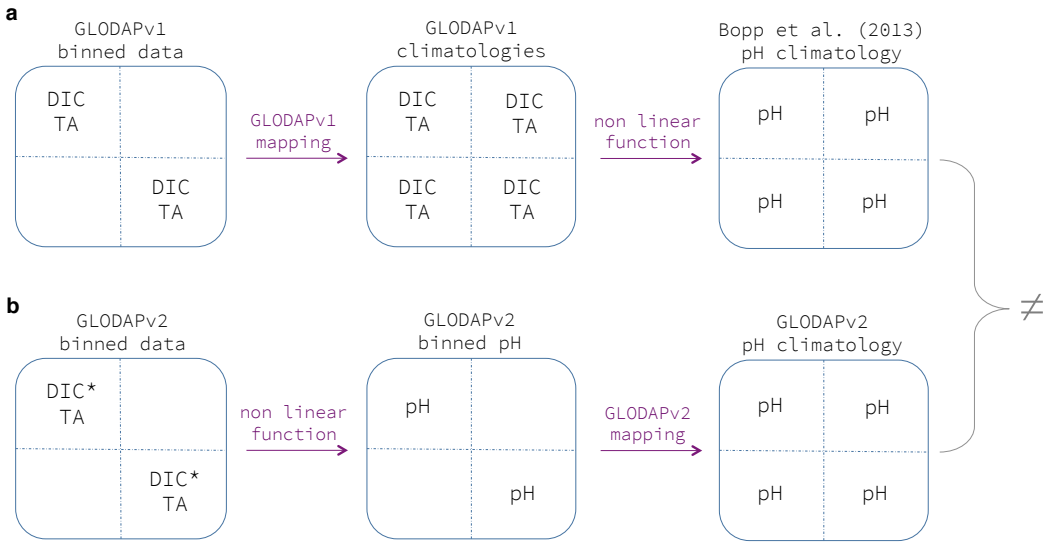

Figure 2.3: Schematic representation of the calculation steps to produce pH climatologies. Maps of pH calculated from the **a**, DIC and TA mapped climatologies from GLODAPv1 as in Bopp et al. (2013), **b**, binned values of DIC normalised in 2002 (referred as DIC* in the figure) and TA as in GLODAPv2.

GLODAP version 2

To improve the climatological products, the GLODAP community extensively developed an updated, unified, bias-corrected product within GLODAPv2 (Table 2.1). GLODAPv2:

- mostly includes measurements taken during the post-CRM era and has bias-corrected some of the pre-CRM era measurements by using more rigorous accuracy methods than what was previously used for the WOCE data (Tanhua et al., 2010; Olsen et al., 2016). Resulting adjustments were particularly large for DIC measurements during the GEOSECS historical measurements (Olsen et al., 2016).

- includes the Mediterranean Sea and Arctic Ocean in the mapped products.

- reduces the risk of converting temporal DIC trends due to increased atmospheric CO_2 concentrations[4] during the observational data coverage (i.e. 1972-2013) into spatial variations, by normalising DIC to 2002 prior to mapping (Lauvset et al., 2016).

- uses an interpolation technique (i.e. Data-Interpolating Variational Analysis (DIVA)) that handles the presence of the seabed and land (Lauvset et al., 2016). This method leads to better results compared to the optimal interpolation method used in GLODAPv1 (Key et al., 2004), particularly near narrow land barriers or across important seabed features (e.g. the mid-Atlantic ridge) (Lauvset et al., 2016).

- provides climatologies for additional parameters (e.g. phosphate, pH). For example, the pH climatology is calculated from binned pairs of TA and normalised DIC (to the year 2002), which then is globally interpolated (Figure 2.3b).

To investigate the potential reasons behind the highly variable spatial patterns in surface pH_{v1} (Figure 2.2b), the mapped pH climatology (hereinafter referred as pH_{v2}; Figure 2.4a) and its associated interpolation error (Figure 2.4b) generated by GLODAPv2 are first described. The spatial pattern in pH_{v2} mostly distinguishes the tropical to the high latitude regions with respective high and low pH values (Figure 2.4a). The interpolation error (Figure 2.4b) is maximal mostly around the coastal western Pacific areas, which therefore

[4]In GLODAPv1, the surface (below 1200 m depth) DIC climatology only used the WOCE data (i.e. the 1990s), but does not provide any year normalisation (Key et al., 2004).

suggests that the spatial distribution in pH_{v2} should be robust except in those coastal areas (Lauvset et al., 2016). Compared to the highly variable spatial patterns in pH_{v1} (Figure 2.2b), pH_{v2} provides a rather zonally uniform distribution[5]. Is the difference in the spatial patterns between pH_{v1} and pH_{v2} due to (1) the change in pH with increasing atmospheric CO_2 (since the two GLODAP versions represent different periods), (2) the choice of the interpolation technique, and/or (3) differences in the calculation steps (pH_{v1} calculates the pH climatology from the DIC and TA climatologies, while pH_{v2} determines the binned pH values before interpolation)? While the hypothesis (1) and (2) cannot be directly tested, the impact of the calculation steps on the pH climatology is investigated.

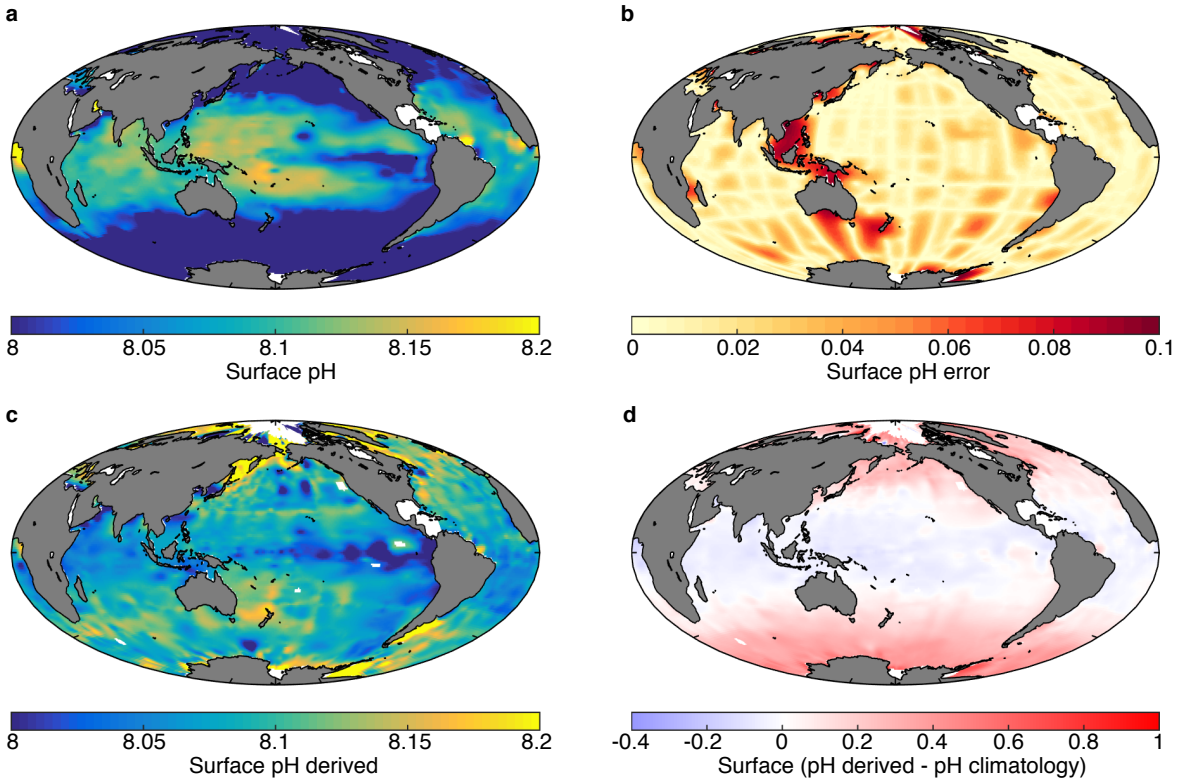

Figure 2.4: The impact of calculation steps in pH climatologies.. **a**, surface pH and its **b**, associated error, provided by GLODAPv2. **c**, surface pH derived from the DIC and TA climatologies from GLODAPv2 using the CO_2SYS Matlab function (Lewis and Wallace, 1998; van Heuven et al., 2011). **d**, difference between (c) and (a). Seawater pH is reported on total scale at 25°C and at 0 dbar (Lauvset et al., 2016; Olsen et al., 2016)

To study the impact of the calculation steps in providing pH climatology, the approach used in Bopp et al. (2013) is reproduced using the GLODAPv2 climatologies. As such, the

[5] pH_{v1} and pH_{v2} are assumed to use identical pH scale.

pH climatology directly provided by GLODAPv2 can be compared to the one derived from the GLODAPv2 DIC and TA climatologies, whose differences would highlight the impact of the ordering in the calculation steps and can therefore indicate if the highly variable spatial pH_{v1} in Bopp et al., (2013) is a relatively consistent signal. The derived surface pH climatology (Figure 2.4c), referred as pH_{v2d}, was specifically calculated from the CO_2SYS Matlab function (version 1.1) (Lewis and Wallace, 1998; Van Heuven et al., 2011) using the salinity, temperature, DIC and TA from GLODAPv2 and the dissociation constants of Lueker et al. (2000) (as in pH_{v2}). Note that the pH_{v2} and pH_{v2d} at constant temperature (25°C) and pressure (0 dbar) is here studied. Interestingly, pH_{v2d} produces a patchier pattern than the smooth pH_{v2}, with much higher pH values (up to almost 1 pH unit) near the coastal areas in the high latitudes and lower values (up to about -0.4 pH unit) in the equatorial regions (Figure 2.4c-d). Assuming that Bopp et al. (2013) used the same dissociation constants, such differences suggest that the variable spatial pattern in pH_{v1} (Figure 2.2b) might be mostly explained by the non-ideal calculation steps rather than being realistic biogeochemical features. As such, the model evaluation that was based on pH_{v1} could present a better variability skill than what was initially estimated by the Bopp et al. (2013) study[6].

Since the publication of Bopp et al. (2013), another climatological pH product has been released (Takahashi et al., 2014), against which surface pH simulated in models can be evaluated (in addition to, or instead of, the climatological pH field provided by GLODAPv2). Unlike GLODAPv2, which calculated pH from pairs of DIC and TA *in-situ* measurements, Takahashi et al. (2014) calculated the climatological pH for the reference year 2005 using the climatologies of surface $pCO_{2\text{-ocean}}$ (Takahashi et al., 2009) and TA. To generate the TA climatology, Takahashi et al. (2014) relied on regional empirical relationships between surface salinity and potential alkalinity, which is defined by the sum of total alkalinity and nitrate concentration and therefore takes into account the effect of biological activities on TA. By using TA measurements rather than DIC measurements[7] (as in GLODAPv2), Takahashi et al. (2014) provides the main advantage, compared to

[6] A model evaluation against the GLODAPv2 pH climatology is not achieved here, as this pH study focuses on highlighting the challenges associated with model evaluation rather than quantifying the pH skill, particularly since pH is not a variable that will be studied in the rest of the thesis.

[7] pH should preferably be defined from $pCO_{2\text{-ocean}}$ and DIC measurements for reasons specified in Takahashi et al. (2014).

GLODAPv2, of providing a pH climatology that is less influenced by seasonal variability in the marine biogeochemical system than if DIC measurements were used. Indeed, since DIC is more sensitive to seasonal variations than TA (c.f. Takahashi et al., 2014 for further discussions), the seasonal observational coverage in DIC measurements should be sufficient to generate appropriate global pH estimates (note that pH estimates should also be based on $pCO_{2\text{-ocean}}$ measurements distributed evenly across the seasons, which is questionable in some poorly sampled basins). However, by developing a methodology based on regional TA and salinity relationships, the Takahashi et al. (2014) method is influenced by the number of measurements available in each region. For instance, due to limited measurements in the highly variable equatorial Pacific (e.g. El Niño events), the TA climatology and hence the pH climatology could not be robustly determined in that region and the corresponding grid cells were therefore left blank; which prevents model evaluations in that region. Also, unlike GLODAPv2 which provides annual climatological fields for 33 depth levels, the Takahashi et al. (2014) product is limited to the surface but provides monthly climatologies. For a complete model evaluation of pH, the scientific community should therefore make use of diversity in the observation-based climatologies (i.e. the GLODAPv2 and Takahashi et al., 2014 products), providing complementary evaluation analyses.

Through the study of Bopp et al. (2013), the pH analysis conducted here provides an example of the various challenges associated with model evaluation and highlights the importance of acknowledging the limitation of an observational-based product to put into perspective the resulting model skill. Note that at the time of their study, Bopp et al. (2013) only had GLODAPv1 available and could not test the importance of pH calculations artefact. Nevertheless, without the quantification of robust uncertainties on interpolated observational products (e.g. uncertainties that would take into account the seasonal bias in the observational coverage in GLODAPv2), model evaluations remain questionable, which therefore limits our understanding of the model's performance and their reliability for future climate predictions.

2.1.3 Investigating summer bias using models

In addition to the spatial and annual irregularities in the DIC and TA measurements (Figure 2.1), the seasonal data distribution is also unbalanced. The total number of winter DIC and

TA measurements is about 4 times lower than in the summer (Figure 2.5). This discrepancy is caused by the challenging winter sea weather conditions, particularly at the high latitudes (Lauvset et al., 2016).

Unless the next decades of global observations become evenly spread across the seasons with potentially the development of biogeochemical floats (e.g. Monteiro et al., 2015), tackling the seasonal bias within the next generation of GLODAP will surely be a challenging task. Here, I investigate in a pure "model-world" the impact of a summer bias on an annual surface climatology, with a focus on the North Atlantic. While it is clear that a summer climatology is going to be different from an annual climatology, I specifically want to illustrate in a model-world the amplitude of this difference and if this difference varies spatially [8].

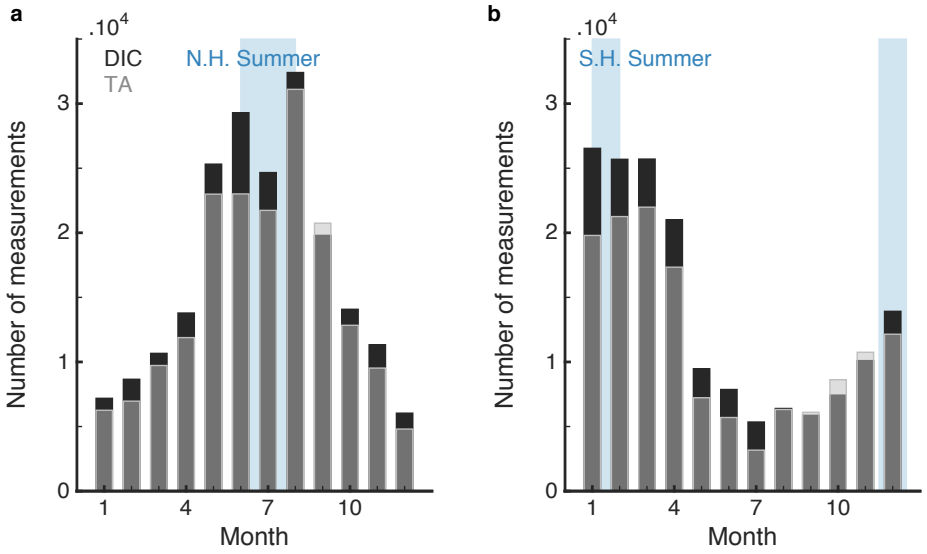

Figure 2.5: Hemispheric monthly distribution of measurements used in GLODAPv2. Total number of DIC (dark grey) and TA (light grey) measurements per month in the **a** North and **b** South hemispheres. The blue box indicates the calendar summer months.

This model-based study uses an ensemble of 15 CMIP5 models (Appendix A.1). For each model, four surface climatologies are calculated: the *summer climatology* of DIC and of TA, and the *annual climatology* of DIC and TA (i.e. the model-truth climatological state) (Table 2.2). The impact of the summer bias to the annual climatology is determined by calculating, for each model, the difference between the two climatologies. A non-zero

[8]The results presented here do not provide, by any means, an attempt to correct for the seasonal bias in GLODAPv2. Such attempt would rely on the fact that the models capture the amplitudes of the real-world biogeochemistry, as well as the phases of the seasonal and natural variability, which, as we will show throughout the rest of the thesis, is not particularly the case.

difference suggests that (1) the annual climatology cannot be captured by using summer data only, and (2) the seasonality, including potentially key winter mechanisms, is important to capture the climatological state.

	DIC	TA
Summer climatology	averaging only summer months of modelled DIC from 1996-2008	averaging only summer months of modelled TA from 1972-2013
Annual climatology	averaging all months of modelled DIC from 1996-2008	averaging all months of modelled TA from 1972-2013

Table 2.2: Description of the four climatologies calculated for each CMIP5 model. In the North Atlantic the chosen summer months are May, June, July and August as they correspond to the months with the highest number of DIC and TA data in GLODAPv2. The periods over which the climatologies were calculated are discussed in Section 2.2.

To understand the impact of the summer months to the annual climatology, the spatial patterns of the DIC and TA climatologies are first presented (Equation 2.1; Figure 2.6 a, d). The model-mean DIC annual climatology presents relatively high values in the central subtropical and subpolar regions, which might be linked to different seasonal processes happening in both regions. In the subtropical region, the relatively high DIC could be explained by the accumulation of carbon at the surface due to strong stratification; while in the subpolar region, the deepening of the mixed layer in winter entrains DIC enriched waters to the surface. The model-mean TA climatology presents two distinctive features that are likely to be linked to precipitation/evaporation distributions, with relatively high TA values in the subtropical gyre and low values in the subpolar.

$$\overline{V_{annual}} = \frac{1}{M}\sum_{m=1}^{M} V_{annual,\,m} \tag{2.1}$$

$$\overline{V_{summer\,bias}} = \frac{1}{M}\sum_{m=1}^{M}\left(V_{summer,\,m} - V_{annual,\,m}\right) \tag{2.2}$$

$$\sigma = \frac{1}{M-1}\sum_{m=1}^{M}\left(V_{summer,\,m} - V_{annual,\,m} - \overline{V_{summer\,bias}}\right)^2 \tag{2.3}$$

where V is the variable of interested (either DIC or TA) and M is the total number of models used (i.e. 15).

The model-based study shows that the DIC summer climatology (Equation 2.2) relatively overestimates the true climatological state in the subtropical gyre (Figure 2.6b) and that this pattern is in good agreement across the models (Figure 2.6c; Equation 2.3). This result is reversed in the subpolar gyre, with an overall underestimation of the summer DIC climatology to the annual climatology and a model disagreement of about 30 to 50 $\mu mol \cdot kg^{-1}$. The larger amplitude difference in the subpolar than in the subtropical gyre suggests that important physical and biogeochemical mechanisms in the subpolar region are not captured by the summer months (e.g. the entrainment of carbon enriched water to the surface occurring during the winter deepening of the mixed layer), while the mechanisms in the subtropical region are overall similar throughout the year. Across the models, the summer DIC climatology adds an overall bias of -2.4 $\pm$ 3.8 $\mu mol \cdot kg^{-1}$ to the annual DIC climatology. On the other hand, the TA model-based analysis shows the climatological state is overall well captured by the summer months only with a bias of 1.0 $\pm$ 3.3 $\mu mol \cdot kg^{-1}$ (Figure 2.6e), also with a good model agreement in the subtropical gyre (Figure 2.6f). Note that those overall biases are relatively small as each regional bias ends up balancing each other. However, if for example the mean error calculations focus on the subpolar region only, the DIC and TA summer biases are -14.5 $\pm$ 11.9 $\mu mol \cdot kg^{-1}$ and -0.1 $\pm$ 9.7 $\mu mol \cdot kg^{-1}$, respectively. If the summer biases identified in this model-based approach are representative of the biases in GLODAPv2 (which we cannot assess), one could therefore imagine how the differences in DIC and TA biases could impact derived products (e.g. anthropogenic CO_2, pH).

While the approach used here illustrates in a purely indicative manner the summer bias in GLODAPv2 using models, one could imagine further use of the models to provide a more comprehensive study on the impact of uneven seasonal measurement distributions on climatologies. For instance, the models' climatologies could be calculated using the same steps as in GLODAPv2, by subsampling the DIC and TA modelled fields at the same locations and times as the observational coverage and interpolating the subsampled fields using the DIVA method. This step could then be repeated by using a virtual observational coverage scheme that adds measurements in winter. By comparing the climatologies between the one based on GLODAPv2 coverage and the one(s) based on a more evenly

distributed measurements, the seasonal bias could be quantified (still in a model-world), or at least be better understood.

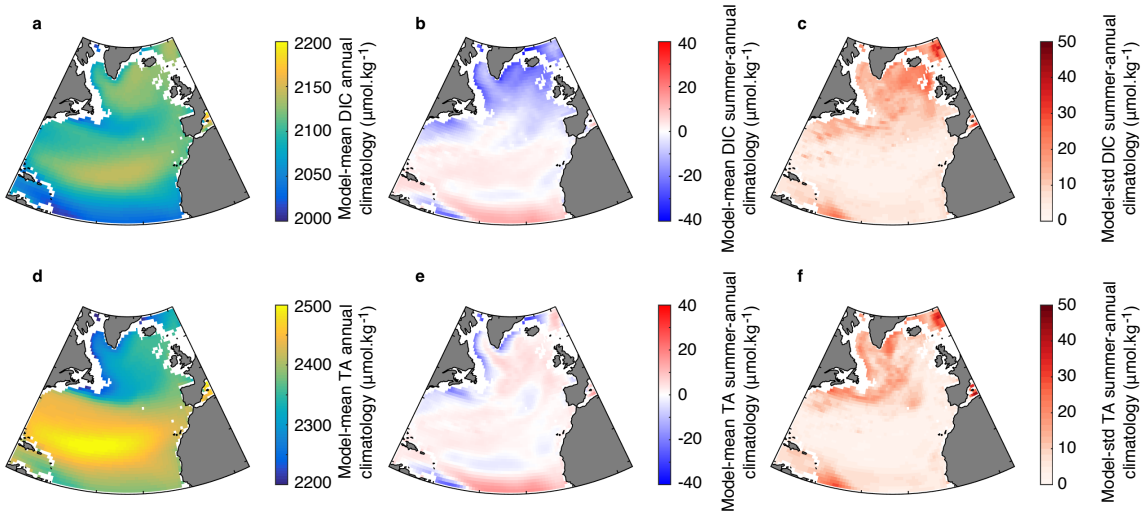

Figure 2.6: Quantification of summer bias on annual climatology in a model-based analysis. Model-mean **a**, DIC and **d**, TA annual climatologies (Table 2.2, Equation 2.1). Model-mean difference between **b**, DIC and **e**, TA summer and annual climatologies (Equation 2.2), where positive/negative values show an overestimation/underestimation of the annual climatology by the summer climatology. Standard deviation across the models of difference between the **c**, DIC and **f**, TA summer and annual climatologies (Equation 2.3).

In summary, synthesising observations into a high-quality product with minimal sources of uncertainties is challenging because (1) measurements present some potential instrumental uncertainties and accuracy errors, (2) mapped products use interpolation techniques whose basin-wide uncertainties are unknown, (3) the data incorporated in those mapped products are irregularly distributed in both space and time, adding some potential biases toward regions and times which are more sampled than others, and (4) derived observational products from interpolated products potentially lead to unrealistic results (through the accumulation of errors due to non-ideal ordering calculation steps). As such, observational products should not be treated as the true state of the system, especially climatologies as they only represent a mean state. Those points should be particularly kept in mind when evaluating the DIC and TA in models.

2.2 Model evaluation

In this section, the North Atlantic biogeochemistry of 15 CMIP5 models (Table A.3) is evaluated against GLODAPv2 (Table A.5). While acknowledging the above challenges, the aim of this study is to (1) get an overall understanding of the models' performance, (2) highlight some potential key model biases, and (3) set a useful background for Chapter 5, which will specifically investigate the impact of those biases on the recent change in the North Atlantic CO_2 uptake. Note that the reasons behind the potential biases identified here is beyond the scope of this chapter, but will be briefly discussed in Chapter 6.

2.2.1 Discussion around data preparation

To provide a relevant evaluation analysis, the model climatologies should be calculated across the same periods as in the observational-based climatologies. However, it is unclear to what degree the GLODAPv2 climatologies are representative of the periods they are intended to represent (Table 2.1). For instance, since the TA measurements were unevenly spread across the 42 year period of available data (i.e. 1972 to 2013), with the majority of the measurements contained within the second half of the period (Figure 2.1), the TA climatology might in fact be more representative of the latter two decades. On the other hand, a TA climatology using model data (which are known at every year and grid cell) from 1972 to 2013 would perfectly represent the mean state of those 42 years. The first challenge in model evaluation is therefore to choose the most appropriate interval; for example choosing between the interval 1972-2013, which is as described in GLODAPv2 but might lead to further interpretation challenges, and a shorter interval that is more representative of the time at which most measurements were made but might capture more internal variability. However, one should ask if TA is expected to change over the 42-year period in the first place. Since large evidence of biological, evaporation/precipitation or circulation changes due to anthropogenic activities have not yet been identified in the North Atlantic through the available observations, TA is not expected to have systematically shifted over the observational period. As such, the interval 1972-2013 was chosen to calculate the TA climatology in the CMIP5 models (Appendix A.1).

The DIC climatology in GLODAPv2 is normalised to 2002 (Lauvset et al., 2016). The nor-

malisation relies on separating the preindustrial component to the anthropogenic component (Lauvset et al., 2016). The anthropogenic carbon in seawater is determined using an exclusively anthropogenic tracer (i.e. the CFC-12 tracer that is gathered within GLODAPv2) and the Transit Time Distribution (TTD) method (Waugh et al., 2006; c.f. Appendix B in Lauvset et al., 2016 for details). For consistency reasons, the DIC normalisation in the CMIP5 models should follow the same normalisation method as the one used in GLODAPv2, but could not be achieved because none of the CMIP5 models simulate the distribution of the CFC-12 tracer (only the GFDL-ESM2M model provides instead the CFC-11 tracer for both the historical and RCP8.5 experiments)[9]. Another way to normalise the models to 2002 is to average the DIC field from the years on either sides, but this leads to a various choice of intervals of different length (e.g. 2001-2003, 2000-2004, 1999-2005). While a short period might predominantly capture the model's natural variability, a longer one might not sensibly represent the model year 2002, especially in a background where the anthropogenic CO_2 rise is not necessarily linear. The other key challenge with model evaluation against a specific date, is that the model year 2002 will almost certainly not be in the same phase of variability as the real-world in 2002. Discrepancies between models and observational-based product might therefore be due to missing mechanisms in models or the models being in a different phase of internal variability, which we cannot assess with the current observational coverage. The interval 1996-2008 was chosen to calculate the DIC climatology in the CMIP5 models (Appendix A.1), as it is centred in 2002 and is a 13-year long interval, which is a similar length as the other climatologies calculated within GLODAPv2 for the variables that are potentially being affected by the anthropogenic change (e.g. pH in Table 2.1).

2.2.2 Evaluating the DIC and TA in the CMIP5 models

The North Atlantic[10] DIC and TA fields of 15 CMIP5 models (Table A.3)[11] are here evaluated against the GLODAPv2 climatologies, at both the near-surface and throughout the water column. To provide a model assessment that captures the various features of the models'

[9]Even if all the CMIP5 models were to provide the distribution of the CFC-12 tracer, the application of the TTD method for each model would have been computationally costly. Indeed, the TTD method follows a series of steps, which are also using the salinity and temperature fields (Waugh et al., 2006).

[10]As a recall from Chapter 1, the North Atlantic is defined from 10°N to 70°N and from -75°E to 5°E.

[11]While the GISS-E2-H-CC and GISS-E2-R-CC models provided DIC, they did not provide TA. For consistency reasons, these two models were removed from the present analysis and as such, all models are evaluated for both their DIC and TA behaviour.

biogeochemistry, complementary model skills' methods are used; from direct spatial and profile comparisons to Taylor diagrams and bias distribution.

Near-surface maps

To put the model evaluation at the near-surface into context, the spatial patterns of the GLODAPv2 DIC and TA are first discussed (in addition to Section 1.2.1), alongside consideration of the interpolation error, which as previously mentioned might be underestimated by 25% (Beckers et al., 2014; Lauvset et al., 2016).

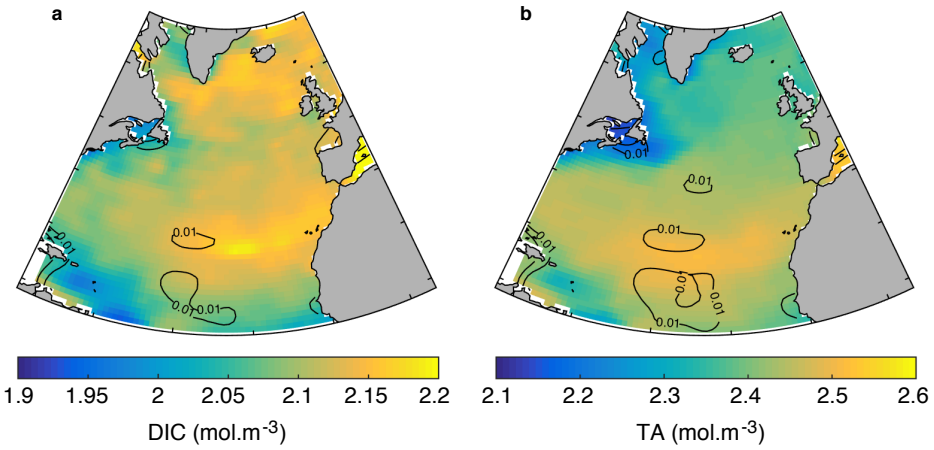

Figure 2.7: The GLODAPv2 DIC and TA climatologies in the North Atlantic. The **a**, DIC and **b**, TA climatologies at 10 m depth, with the corresponding interpolation error indicated by the contour lines.

Across the North Atlantic, the near-surface GLODAPv2 DIC climatology is approximately uniform (Figure 2.7a). An enhancement in near-surface DIC is observed in the eastern subtropical gyre (Figure 2.7a), which could be explained by the strong stratification preventing the dilution of enriched carbon surface waters with waters below the thermocline, and/or by the transport of upwelled DIC-enriched waters from the African coast to the open ocean, but could also result from an interpolation artefact as indicated by the interpolation error in that area. The near-surface TA climatology presents higher values in the subtropical gyre than in the subpolar region (Figure 2.7b), which seems to follow the evaporation/precipitation regimes (Section 1.2.1). Compared to the rest of the basin, relatively low values in both DIC and TA are also noticed near the North American coast and in the Caribbean sea (Figure 2.7), which could be due to specific oceanic circulation

(e.g. the input of fresh Arctic waters, with relatively low DIC and TA, from the Labrador current that feeds the North Atlantic region and river runoff into the Gulf of Mexico) or interpolation artefacts, shown by the relatively high interpolation error near those areas.

The first approach to evaluate the models' near-surface DIC and TA is to directly compare those fields against the GLODAPv2 equivalents. Specifically for each CMIP5 model, the near-surface DIC and TA fields are respectively subtracted by the near-surface GLODAPv2 DIC and TA, forming a residual map for each biogeochemical field. As such, the residual spatial patterns indicate regions where the model overestimates (i.e. positive residuals) or underestimates (i.e. negative residuals) the observation-based biogeochemical climatologies. The main discrepancies between the models and the GLODAPv2 climatologies can describe important model biases but can also arise from the differences between how the model and observation-based climatologies were produced, particularly due to the various sources of uncertainties behind GLODAPv2 (e.g. the summer bias present in the GLODAPv2 fields which does not exist in the models climatologies as complete annual fields were used) and/or the chosen periods for the models' climatologies (Section 2.2.1). The resulting comparison might therefore highlight differences in the climatology production rather than poor model skills. Nevertheless, since the DIC and TA residuals (Figures 2.8, 2.9) are about 10 times higher than the amplitude of the interpolation error (Figure 2.7), it suggests that the residual signals are potentially independent from the interpolation error.

The near-surface DIC residuals present relatively different patterns across the 15 CMIP5 models (Figure 2.8). While most models overestimate the near-surface DIC field across the North Atlantic (e.g. GFLD-ESM2G, IPSL-CM5A-LR, MPI-ESM-MR and particularly NorESM1-ME), others present an overestimation in the subtropical region and an underestimation in the subpolar region (especially MRI-ESM1) (Figure 2.8). Other substantial patterns of underestimated DIC are found in a few models (e.g. CanESM2, MIROC-CHEM, MPI-ESM-LR) around the Gulf Stream and at the boundary between the subtropical and subpolar gyres (Figure 2.8). Since the near-surface DIC in GLODAPv2 does not provide a distinctive feature at the gyre boundary (i.e. the DIC field is rather uniform across the basin),

the localised DIC underestimations in those models therefore arise from the existence of an unrealistic feature at the gyre boundary. In the concerned CMIP5 models (e.g. CanESM2, MIROC-CHEM, MPI-ESM-LR), the lower DIC at the gyre boundary compared to the rest of the basin could be explained by the strong cooling that surface waters experience when moving from the subtropical to subpolar basins that perturbs the equilibrium state of the two gyres, or by the fact that the low-DIC Arctic waters are being recirculated from the Labrador Sea to the subpolar gyre via the North Atlantic Current. Finally, the CESM1-BGC model seems to display the lowest amplitudes in the residuals across the basin (Figure 2.8), suggesting that this model qualitatively presents the highest DIC skill relative to GLODAPv2 of all models used here.

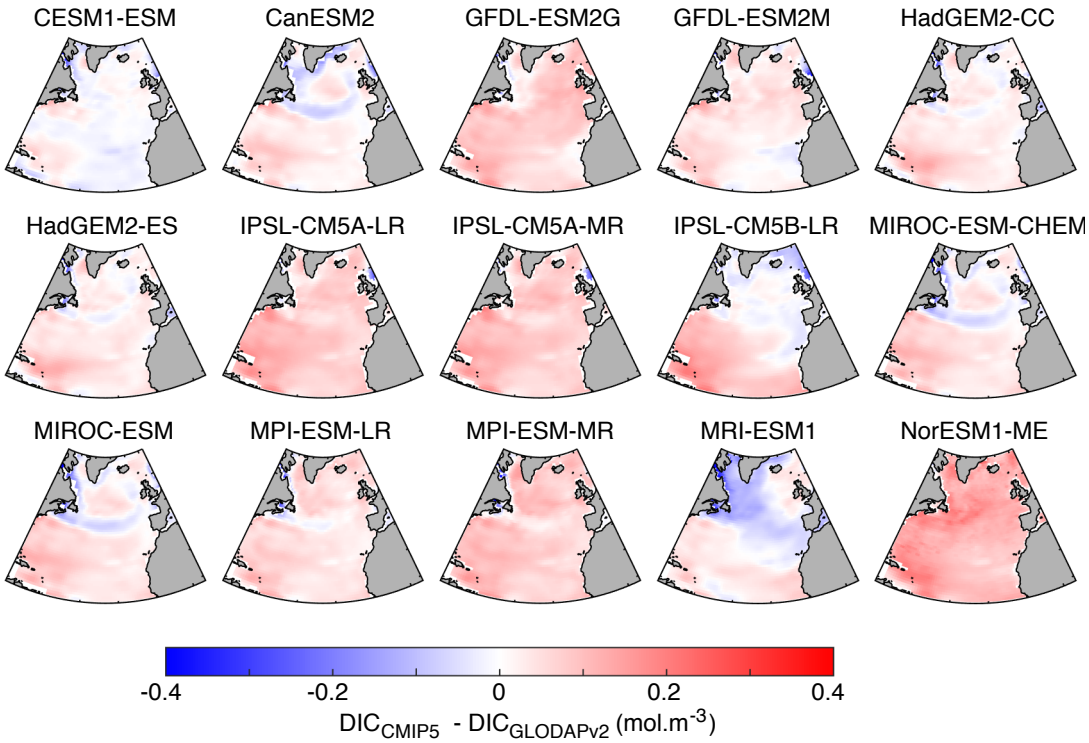

Figure 2.8: Near-surface DIC residuals in the North Atlantic. Difference between each CMIP5 model and GLODAPv2 DIC climatologies (i.e. the residuals) at 10 m depth.

Compared to the near-surface DIC residual patterns, the TA residuals generally present a higher spatial variability of higher amplitude (Figure 2.9). Apart from NorESM1-ME which distinctively overestimates the GLODAPv2 TA across the whole North Atlantic basin, the CMIP5 models show an overestimation in the subtropical region and an underestimation in the subpolar region (particularly in the MRI-ESM1 and IPSL-CM5B-LR models) (Figure

2.9). A particular plume of negative residuals (i.e. an underestimation of TA by the models) is visible in most models also around the Gulf Stream area and at the gyres boundary (e.g. HadGEM2-CC, IPSL-CM5A-LR, MPI-ESM-MR), which as mentioned could be due to specific oceanic patterns in the models (e.g. the intense physical change between the gyres perturb in return the biogeochemical properties in the zone of transition) or strong seasonal changes in that region that might have been missed in the GLODAPv2 product. A local feature is also noticed along the North American coast (i.e. positive TA residual signal), which could arise from the input of freshwater from river inputs (low alkalinity) missed by the models or the relatively high interpolation errors issues in GLODAPv2 in that area (Figure 2.7b).

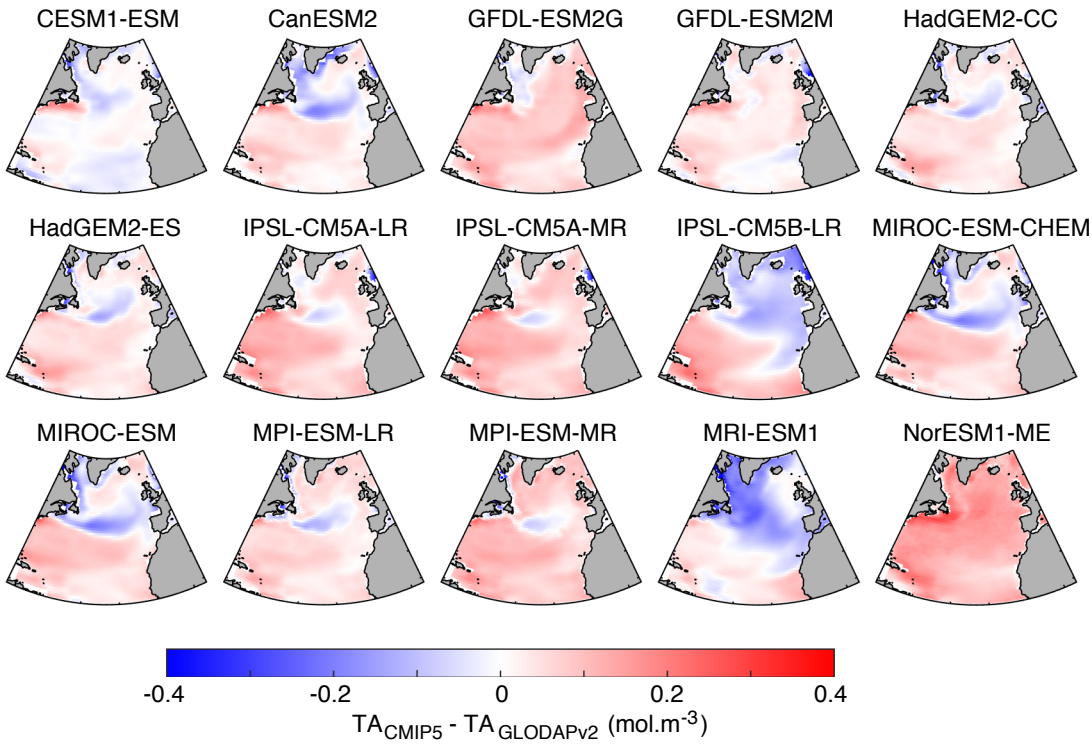

Figure 2.9: Near-surface TA residuals in the North Atlantic. Difference between each CMIP5 model and GLODAPv2 TA climatologies (i.e. the residuals) at 10 m depth.

Vertical profiles

The models' DIC and TA are also evaluated at depth first through a basic model-observation qualitative comparison (Figure 2.10). The MRI-ESM1 model presents unrealistic profiles, which are linked to issues with the regridding step (issues were hence spotted during the current analysis) (Appendix A.1). Indeed, the MRI-ESM1 model used a land mask for the

4-dimension DIC and TA fields that was filled with zeros instead of a common "not-a-number" mask. As such, the regridding in coastal areas must have been impacted when interpolating oceanic DIC and TA values with zero values on land, explaining the model's unrealistic averaged profiles (note that on the regridded fields, the remaining zeros on land were removed before calculating the averaged profiles). While further work could have been done to get around this issue (e.g. removing coastal areas in the regridded product, re-processing the model data after the zeros in the land mask removed), I highlight here some of the many difficulties when dealing with large datasets of different sources, which do not always follow the CMIP5's consensus. For completeness reasons, the evaluation of the MRI-ESM1 model is included in the subsequent figures but will not be discussed in the text.

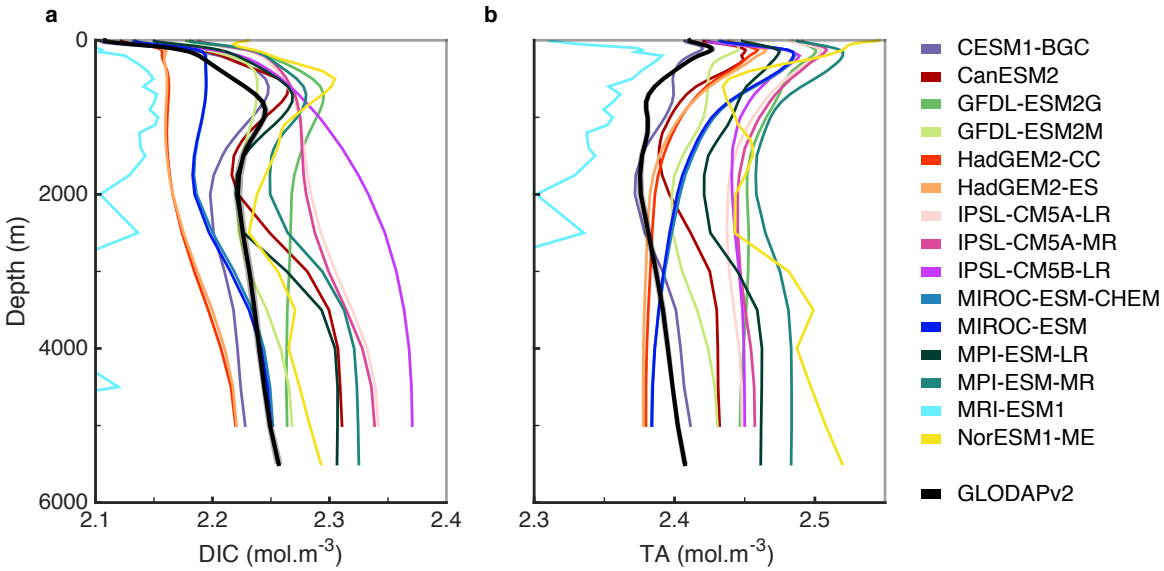

Figure 2.10: DIC and TA model evaluation at depth. **a**, DIC and **b**, TA area-weighted mean profile in the North Atlantic from the 15 CMIP5 models, and from GLODAPv2 (thick black). Note that the area-weighted mean interpolation error is displayed around the observation-based climatology profiles but is indiscernible from the mean profile, and that the x-axis limits purposely crop the MRI-ESM1 profiles because of their large divergence at depth with the other models, which would have led to difficult visualisation for the other models otherwise.

The CMIP5 profiles are distinctive in two aspects: while the models' DIC profiles are evenly distributed around the GLODAPv2 profile, the models' TA profiles systematically overestimate the GLODAPv2 profile over almost the entire water column (Figure 2.10). The shapes of the models' DIC profiles also differ from the shape of the GLODAPv2 profile, with for instance a shallower maxima in the intermediate waters (Figure 2.10), which could lead

to further impact on the models' surface DIC during relatively shallow mixing events. Nevertheless, a good model agreement with the GLODAPv2 profiles might not necessarily arise from a good biogeochemical model description, but could be explained by a relatively short spin-up duration or a type of spin-up procedure that has prevented the model to substantially drift away from the initialisation conditions, which are often set from observational-based products (Séférian et al., 2016).

Taylor diagrams

To synthesise the qualitative model-observation comparisons at the near-surface and at depth into specific quantitative metrics, Taylor diagrams on the DIC and TA fields are studied (Taylor et al., 2001). As explained in Chapter 1 (Section 1.5), Taylor diagrams efficiently compare the skill of different models relative to the observational product within a single visual. Specifically, the Taylor diagram displays through a polar coordinate system, the standard deviation along the radial dimension and the correlation between the model and observation fields through the polar angle (Taylor, 2001; Jolliff et al., 2009). Isolines of unbiased RMSE are also indicated (green lines on Figure 2.11) (Friedrichs et al., 2009). As such, the observational reference point has a polar angle of 0° (i.e. correlation of 1) and is on the 0 mol·m^{-3} RMSE isoline and here, the standard deviation of near-surface DIC and TA in GLODAPv2 is 0.05 mol·m^{-3} and 0.04 mol·m^{-3}, respectively (Figure 2.11). Taylor diagrams are therefore read by evaluating the distance of each model point from the observational reference point, which allows to easily identify clusters across models. Here, Taylor diagrams have been computed using for each model, all available grid cells of the North Atlantic DIC and TA fields (including near-surface and vertical data).

The DIC Taylor diagram presents two model clusters: one with relatively high correlations (between 0.80 and 0.90), low RMSE (between 0.02 and 0.03 mol·m^{-3}) and standard deviations that broadly capture the variance of the GLODAPv2 DIC; and a second cluster with low correlations (between 0.50 and 0.75), high RMSE (around 0.04 mol·m^{-3}) and much wider variability than the GLODAPv2 DIC (Figure 2.11a). The TA Taylor diagram does not present any particular model clustering, but rather that the models systematically overestimate the variance of GLODAPv2 (Figure 2.11b). Nevertheless, the

models that had the highest correlation with the observation-based DIC (models from the first cluster) also present the highest correlation (between 0.8 and 0.9) with the observation-based TA (except for the MRI-ESM1 model, which for the reasons mentioned above is discarded from the discussion). Among the CMIP5 models, the GFDL-ESM2M model is identified with the best DIC and TA skill (highest correlation with the GLODAPv2 fields, smallest unbiased RMSE and closest to the reference point), and the IPSL-CM5A-LR model is identified as the poorest skilled model (Figure 2.11). Interestingly, the other two models from the IPSL modelling group (i.e. IPSL-CM5A-LR and IPSL-CM5A-MR) provide a much better skill than the IPSL-CM5B-LR model, which was at first not particularly expected as the models used the same biogeochemical model (Table A.1). While the different atmospheric configurations between the IPSL-CM5A and IPSL-CM5B versions might have had an impact on the ocean circulation and hence on the model skill of DIC and TA, particularly through different representations of the boundary layer (Dufresne et al., 2013), the difference in the spin-up protocols might in fact have had a major role in setting up the background DIC and TA (Séférian et al., 2016).

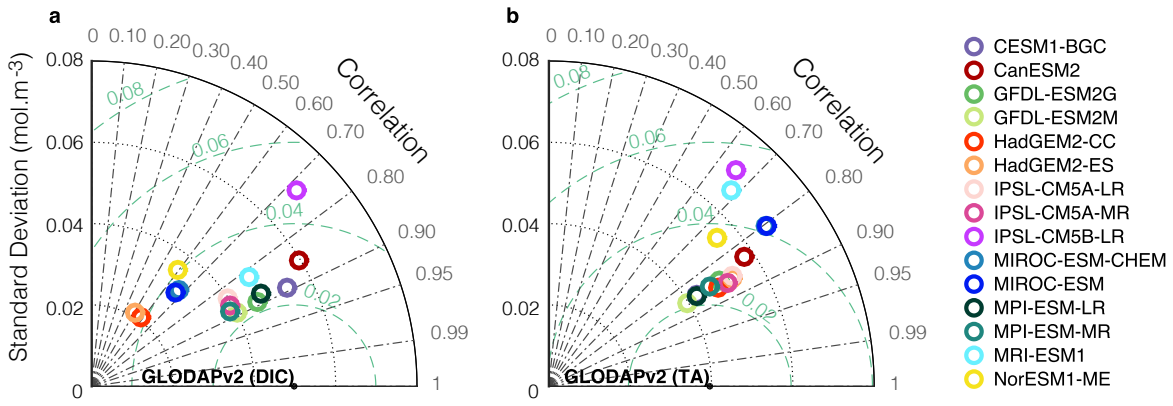

Figure 2.11: DIC and TA Taylor diagrams. **a**, DIC, and **b**, TA Taylor diagrams (Taylor, 2001) based on the all available grid cells of all depth. The green lines indicate the unbiased RMSE (in mol·m^{-3}).

Normalised bias

While Taylor diagrams are powerful and concise, they do not contain any bias information (Jolliff et al., 2009) (Section 1.5). On the other hand, the near-surface residual maps (Figure 2.9) and the vertical profiles (Figure 2.10) indicate regions/models with low or strong biases

but are ultimately qualitative approaches and lead to challenging model inter-comparisons. To quantify the models' DIC and TA bias relative to the GLODAPv2 climatologies, the normalised bias B is here calculated for each model and at each depth:

$$B_d = \left\{ \frac{\overline{V_{CMIP5}} - \overline{V_{GLODAPv2}}}{\sigma_{V_{GLODAPv2}}} \right\}_d \tag{2.4}$$

where at depth d, $\overline{V_{CMIP5}}$ is the area-weighted mean of a CMIP5 model for the field V (i.e. either DIC or TA), $\overline{V_{GLODAPv2}}$ the equivalent for GLODAPv2 and $\sigma_{V_{GLODAPv2}}$ is the standard deviation from the area-weighted mean of the GLODAPv2 field (Jolliff et al., 2009). The normalised bias is therefore a dimensionless variable, which allows to compare the range of biases between the DIC and TA fields.

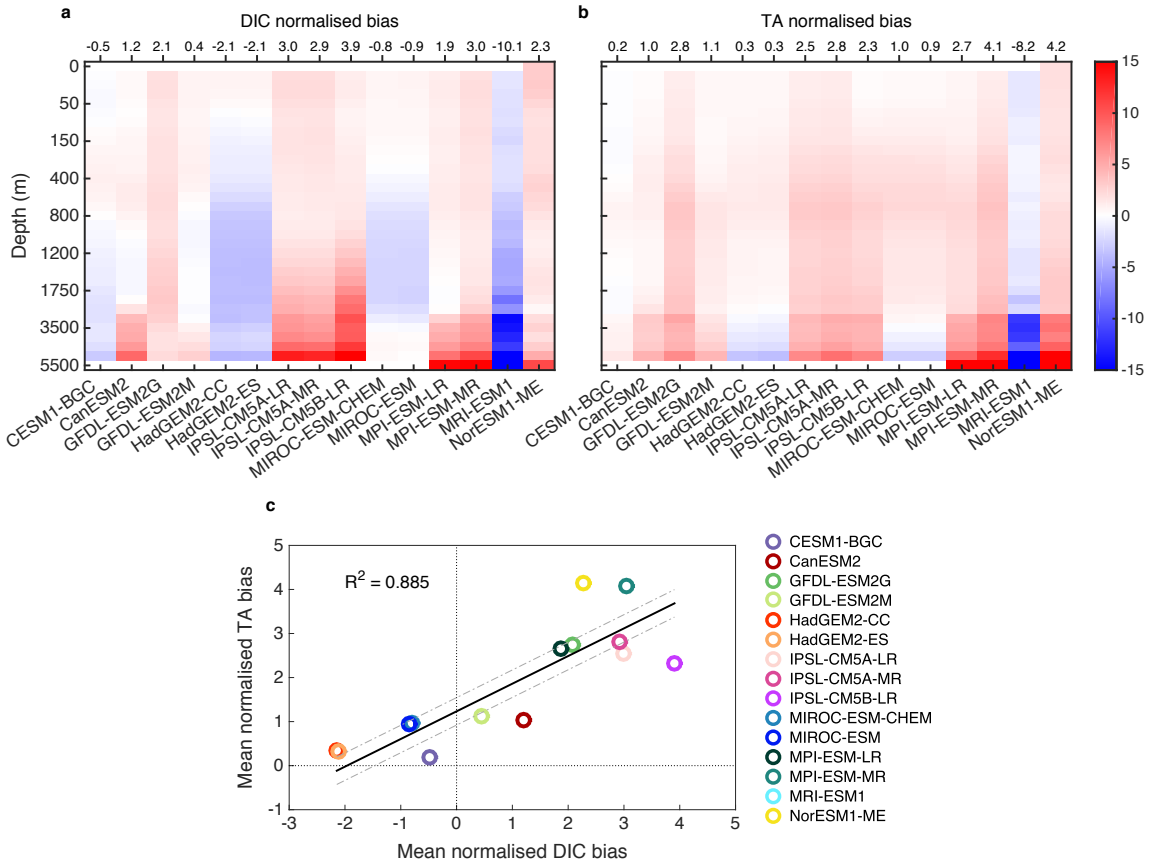

Figure 2.12: Normalised DIC and TA biases. Vertical profiles of the normalised **a**, DIC and **b**, TA biases for each model (Equation 2.4). The number on top of **a** and **b** indicates the averaged normalised bias value throughout the water column, which are also displayed as a scatter plot in **c** (except for MRI-ESM1). In **c**, the black line indicates the linear fit and the dashed lines correspond to the 95% confidence intervals.

The normalised bias study presents coherent results with the above analyses (i.e. a range of positive and negative biases for the DIC field, an overall positive bias for the TA field) but also highlights new features (Figure 2.12a,b). For instance, the GFDL-ESM2M and IPSL-CM5A-LR models, which were identified from the Taylor diagrams with respectively the best and poorest model skills, actually present an absolute normalised bias in TA that respectively ranks the 7th and 9th position out of the 14 models (ignoring MRI-ESM1; Figure 2.12a,b); the lowest absolute normalised TA bias being CESM1-BGC (which can be seen as closely following the GLODAPv2 TA profile on Figure 2.10). In the case of the GFDL-ESM2M model, the different conclusions between the Taylor diagrams and the bias analysis on the TA field is explained by the fact the model profile has a similar shape to the GLODAPv2 TA profile, but whose amplitude presents an offset (Figure 2.10). The amplitude of the model biases in DIC and TA fields also generally increase with depth, especially below 2,500 m depth, which for example suggests that models are missing key processes of the biological pump at depth or that they poorly simulate the overturning oceanic circulation (further discussions in Chapter 6), or that the GLODAPv2 interpolation uncertainty is particularly underestimated at depth (potentially due to fewer available data and bathymetry interactions).

Interestingly, models that have a relatively low/high mean normalised TA bias also tend to have a relatively low/high mean normalised DIC bias, respectively. A positive correlation (R^2=0.885) between the two mean normalised biases is found, as well as a statistically significant linear gradient of 0.63 ± 0.07, which indicates that the models broadly present a larger bias in TA than in DIC (Figure 2.12c). While the causality between the DIC and TA biases cannot be tested, one may suggest that the relatively high TA biases potentially drives the biases in DIC. Indeed, an initial positive model bias in TA (e.g. potentially from model drift during the spin-up period) would result in additional DIC being formed during the model's spin-up through the suppression of surface $fCO_{2\text{-ocean}}$ taken up from the atmosphere. As we will further discuss in Chapter 5, the TA and DIC biases in the CMIP5 models are likely to be responsible for the systematic model underestimation of the recent change in the North Atlantic surface $fCO_{2\text{-ocean}}$. If the TA biases are indeed the major driver for the biases in other marine biogeochemical variables, the reason(s) behind the TA biases in the CMIP5 models should be identified and tackled by the modelling community. However, identifying

the exact reason(s) for the TA biases in the models is challenging, particularly because each CMIP5 model have their own complexity, spinup protocols, representation of the carbon cycle and ocean circulation. While physical-driven mechanisms that could be responsible for the TA biases (i.e. errors in freshwater inputs and errors in the ocean circulation) will be discussed in Chapter 6, here we discuss the role of biology in setting appropriate TA profiles. In the real-world, the vertical distribution in TA is influenced by biology through the soft-tissue pump and the carbonate pump (c.f. Section 1.2.2). While the soft-tissue pump induces a decrease in TA with depth due to the remineralisation of organic matter, the carbonate pump leads to an increase in TA with depth due to the dissolution of calcium carbonate $CaCO_3$ (c.f. Figure 1.7). As such, the relatively constant TA between 1,000 and 3,000 m depth might be explained by compensating effects from the two biological pumps, and the increase in TA below 3,000 m depth is likely to be explained by the dominance of the carbonate pump over the soft-tissue pump. Interestingly, the carbonate pump is represented in the CMIP5 models only by the calcite mineral species, while in the real-world, the aragonite mineral species also exists (except for the two GFDL models which simulate both calcite and aragonite in their ocean biogeochemical model). The merging of the calcium carbonate species in the models could lead to substantial errors in the vertical distribution of TA, that might have accumulated during the spinup period. The fact that the GFDL models simulate aragonite and provide a relatively good TA and DIC variability (especially for the GFDL-ESM2M model, which has a z-coordinate physical ocean model) suggest that the modelling community should focus on further developing the carbonate pump in ocean biogeochemical models to potentially improve the TA and hence DIC distributions in the oceans. Indeed, the inclusion of aragonite in the carbonate pump would not necessarily improve the TA representations in the models if the models' ocean circulation remains inappropriate and/or the spinup is not adequately tuned. However, to provide adequate tuning during the spinup phase and accordingly evaluate the models, the amount of calcification occurring in the oceans should be better understood and quantified by the observational community.

In summary, the extensive model evaluation of the DIC and TA fields of 15 CMIP5 models against the GLODAPv2 climatologies has allowed the identification and discussion of

three main points: (1) as each evaluation method provides their strength and limitations, the use of complementary techniques is necessary to improve our understanding of the models' performance, (2) the CMIP5 models seem to broadly present a systematic positive TA bias throughout the water column, suggesting that a common modelling feature or setup is largely responsible this positive bias (e.g. models drifting from the initial conditions during the spin-up phase), and (3), due to the various spin-up durations, protocols and initial conditions used across the CMIP5 models (Séférian et al., 2016), the resulting DIC and TA features might therefore illustrate differences in the spin-up setup rather than in the models' performance, leading to further challenges when understanding the reasons behind potential model biases and comparing the models' response to increasing atmospheric CO_2 concentrations.

2.3 Summary

Over the past decades, our understanding of the carbon system has expanded thanks to the relentless international efforts to develop, improve and maintain measurements across the globe, and also to the extensive work from the GLODAP community to produce a high-quality synthesised product. While such products are particularly useful for model evaluation studies, the often overlooked uncertainties (e.g. measurement accuracy, interpolation error) can lead to challenges when interpreting model skill. A case-study analysis highlighted the importance of acknowledging the limitations of an observational-based product, particularly on derived observational properties (e.g. pH derived from DIC and TA climatologies), in order to put the models' performance into perspective. Within this context, the North Atlantic DIC and TA fields of 15 CMIP5 models were evaluated against the GLODAPv2 climatologies. Overall, the models DIC and TA seem to present a mean normalised biases of 1.0 ± 2.0 and 1.9 ± 1.3 through the water column, respectively. The model skill analysis particularly suggests a systematic positive bias in the models' TA, which could be linked to model drift reasons during the spin-up phase. Such results could have important repercussions for setting appropriate change in oceanic carbon uptake, particularly through modifications of the CO_2 buffering capacity (investigated in Chapter 6). While GLODAPv2 allows us to understand the broad features of the marine carbon system, examination of the climatological state alone does not allow one to investigate any temporal change and/or the biogeochemical

modes of variability. To investigate the potential impact of the identified DIC and TA biases on the time-varying air-sea CO_2 flux (Chapter 5), the models' surface ocean fugacity, which is a crucial variable for setting the direction and magnitude of the flux, is evaluated in the following chapter.

Chapter 3

Evaluating the sea surface fugacity of CO_2 in Earth System Models

This chapter aims to evaluate the North Atlantic sea surface fugacity of CO_2 ($fCO_{2\text{-ocean}}$) in the CMIP5 models against the Surface Ocean CO_2 Atlas (SOCAT) version 4 gridded product, particularly highlighting the strengths and limitations of this observational product for quantifying the basin-wide CO_2 oceanic uptake and for model evaluation purposes.

Unlike GLODAP, which produces mapped climatologies of the ocean biogeochemistry (Chapter 2) or provides non-gridded bottle data, SOCAT provides surface observational $fCO_{2\text{-ocean}}$ data as original observations of each data set, as well as a gridded product of monthly mean $fCO_{2\text{-ocean}}$ in 1° latitude by 1° longitude grid cells where observations exist, and hence provides a time-varying but not interpolated product. As such, the monthly gridded SOCATv4 product makes it a useful product for the evaluation of temporal variability within the surface CO_2 system simulated by models. Here, the monthly gridded SOCATv4 product is presented with a focus on the spatial and temporal data coverage in the North Atlantic and how such coverage can impact $fCO_{2\text{-ocean}}$ trend studies (Section 3.1). With this context, the evaluation of the surface $fCO_{2\text{-ocean}}$ from 20 ESMs is carried out (Section 3.2). By highlighting the challenges associated with the spatially and temporally incomplete nature of the $fCO_{2\text{-ocean}}$ observations coverage, this chapter will set an essential background for the following two chapters.

3.1 The SOCATv4 gridded product: strengths and limitations

The ocean fugacity of CO_2 (i.e. $fCO_{2\text{-ocean}}$) is one of the four parameters defining the carbonate system and plays an important role in governing the air-sea CO_2 flux (Chapter 1; Takahashi et al., 2002; Arruda et al., 2015). To understand and quantify the modes of variability within the oceanic CO_2 sink, as well as its regional patterns, it is therefore necessary to maintain surface ocean CO_2 observations. Over the past decades, surface ocean CO_2 observations have been collected through the development of autonomous instrumentation on board research vessels and VOS (i.e. commercial ships), on moorings and more recently on autonomous vehicles (Monteiro et al., 2015). To gather the substantial number of measurements collected through such broad international efforts, the SOCAT database was created and whose version 4 records from 1.9 million measurements from the 1990s to 8.4 million from the 2000s (Pfeil et al., 2013; Bakker et al., 2016). The monthly gridded version of SOCAT is a freely available product (http://www.socat.info/) that gathers uniformly quality-controlled $fCO_{2\text{-ocean}}$ observations[1] since 1970 into a regular 1°x1° monthly grid (Pfeil et al., 2013). The active SOCAT community regularly updates its product with newly available data, leading to the creation over the past six years of five successive versions. Here, as well as in the rest of the thesis, the SOCAT version 4 monthly gridded product is used, hereinafter referred as "SOCATv4" (Sabine et al., 2013; Bakker et al., 2016; Table A.5). Note that independently from SOCAT, the Lamont-Doherty Earth Observatory (LDEO) database also provides a synthesis of surface CO_2 observations covering the period from 1957 to 2016 (Takahashi et al., 2017). Specifically, the LDEO database provides, in a non-gridded format, global $pCO_{2\text{-ocean}}$ observations available for each date with measurements (taken from semi-continuous underway $pCO_{2\text{-ocean}}$ systems and deduced from discrete bottled data) (Takahashi et al., 2017). While the SOCAT and LDEO products mostly contain similar original sources of data, their treatment and quality control differ and as such, are treated as independent products (Bakker et al., 2016). In the context of this thesis, the focus is made on the SOCATv4 product only.

This section aims to give an insight of the SOCATv4 monthly gridded product, starting with

[1] Since $fCO_{2\text{-ocean}}$ is not a directly measured quantity, I refer to this quantity as "$fCO_{2\text{-ocean}}$ observations" and not "$fCO_{2\text{-ocean}}$ measurements".

a presentation of the spatio-temporal distribution of the gridded $fCO_{2\text{-ocean}}$ data at the global scale and in the North Atlantic (Section 3.1.1), followed by a discussion around some of the sources of uncertainties behind such product (Section 3.1.2), and ending with a qualitative study on the impact of the gaps in the spatial and temporal coverage on the $fCO_{2\text{-ocean}}$ trends in the North Atlantic (Section 3.1.3).

3.1.1 Presentation of the SOCATv4 gridded product

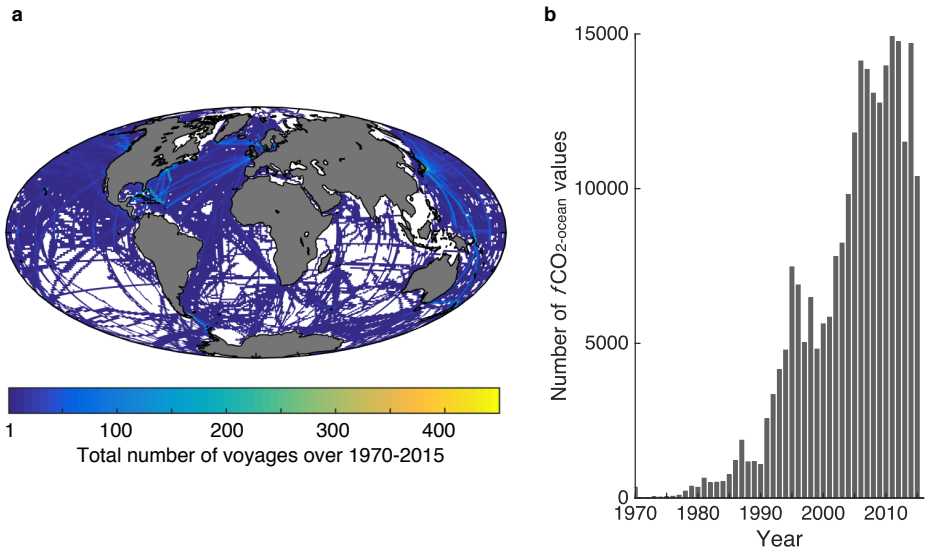

Figure 3.1: Global spatial and temporal description of SOCATv4. **a**, Spatial distribution of the total number of voyages taking place between 1970 and 2015. **b**, Total number of global $fCO_{2\text{-ocean}}$ values per year.

SOCATv4 provides $fCO_{2\text{-ocean}}$ monthly voyage weighted mean values on a 1°x1° grid with no interpolation (Sabine et al., 2013; Bakker et al., 2014). The voyage weighted mean is calculated by, for each grid cell and month with data, first averaging all the $fCO_{2\text{-ocean}}$ observations recorded by each voyage separately, and then averaging the $fCO_{2\text{-ocean}}$ observations across the various voyages (Sabine et al., 2013). The voyage weighted mean therefore reduces the bias introduced by voyages that have high temporal resolution measurements (Sabine et al., 2013). The SOCATv4 monthly gridded product contains data from 1970 to 2015 and contains 239,473 grid cells with $fCO_{2\text{-ocean}}$ values, which were deduced from about 18.5 million $fCO_{2\text{-ocean}}$ observations over that interval. Hereinafter, the term "$fCO_{2\text{-ocean}}$ values" refers to the SOCATv4 gridded values, calculated from the $fCO_{2\text{-ocean}}$ observations that were submitted to the SOCAT database. SOCATv4 also holds

18 grid cells with $fCO_{2\text{-ocean}}$ values that are higher than 1,000 µatm (specifically 18 values between 1,021 and 4,310 µatm), which are located in estuaries areas (mostly in the North Atlantic coastlines) and are likely to describe periodic impacts of river runoff. Those 18 $fCO_{2\text{-ocean}}$ values were therefore qualified as outliers (although realistic) and were removed for the rest of this chapter.

Over the past 45 years, the number of global $fCO_{2\text{-ocean}}$ values has notably increased, and has maintained high number of values (i.e. above 100,000) since 2005 (Bakker et al., 2016) (Figure 3.1b). Note that the year 2015 contains slightly fewer $fCO_{2\text{-ocean}}$ values than previous years, which is expected due to not-yet submitted data at the time of the SOCATv4 release. While the Southern Hemisphere reports less than 30% of the total number of $fCO_{2\text{-ocean}}$ values and presents substantial gaps in the South Pacific and Indian Ocean, the intense commercial shipping routes in the North Pacific (particularly in the western part) and North Atlantic set these basins as the most sampled regions on the planet (Figure 3.1a).

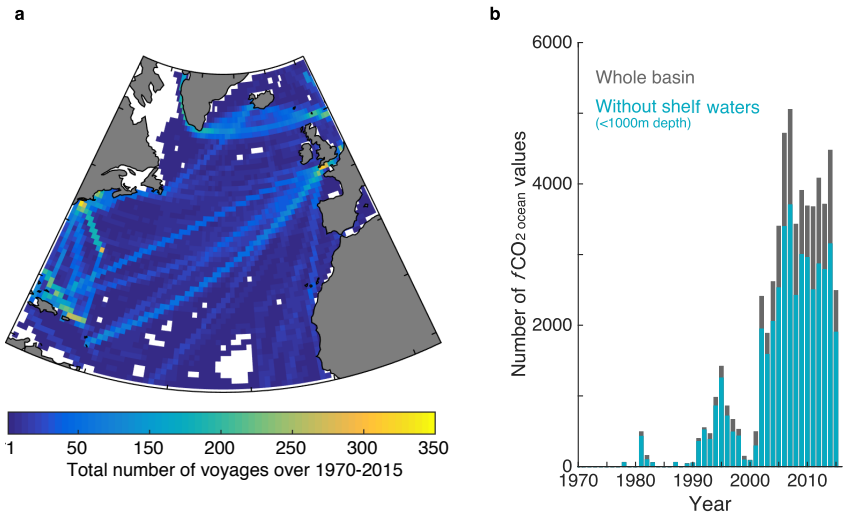

Figure 3.2: Spatial and temporal description of SOCATv4 in the North Atlantic. **a**, Spatial distribution of the total number of voyages between 1970 and 2015. **b**, Total number of $fCO_{2\text{-ocean}}$ values per year for the whole basin (grey bar) and after removing the shelves waters that are below 1,000 m depth (blue bars).

The North Atlantic[2] holds 57,006 $fCO_{2\text{-ocean}}$ values, with more than 50% contained within the period 2007-2014, and about 75% being in the open ocean (i.e. waters deeper that 1,000 m depth; Figure 3.2b). The majority of the data in the North Atlantic was collected by sustained shipping routes, particularly the UK to Caribbean line, and by research

[2]A reminder that the North Atlantic is defined from 10°N to 70°N and from -75°E to 5°E.

programmes in the subpolar region (Figure 3.2a). Despite the international efforts to extend the observational coverage, some areas still remain empty of any measurements (e.g. the North Labrador Sea, southern subtropical gyre) (Figure 3.2a).

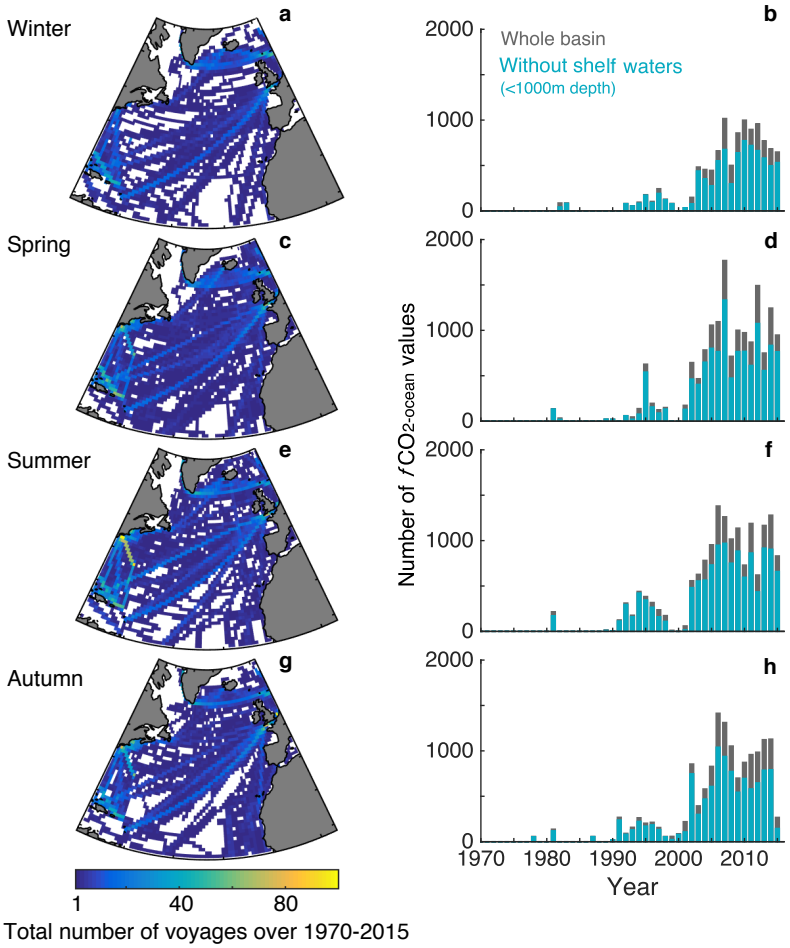

Figure 3.3: Seasonal description of SOCATv4 in the North Atlantic. **left**, Spatial distribution of the total number of voyages during **a**, winter (DJF), **c**, spring (MAM), *e*, summer (JJA) and *g*, autumn (SON) from 1970 to 2015. **right**, Total number of $fCO_{2\text{-ocean}}$ values per year per season (i.e. **b**, winter, **d**, spring, **f**, summer, **h**, autumn) for the whole basin (grey bar) and after removing the shelves waters that are below 1,000 m depth (blue bars).

In the North Atlantic, the SOCATv4 distribution also varies across seasons (Figure 3.3). While summer is the season with the highest number of $fCO_{2\text{-ocean}}$ values across the period 1970-2015 (a total of 16,044 values, against 15,568, 14,546 and 10,730 for spring, autumn and winter, respectively), spring is actually the season that provides the best spatial coverage during that period (with 2,961 filled grid cells, against 2,902, 2,754 and 2,440 for summer, autumn and winter, respectively). As such, the SOCATv4 product seems to present a seasonal bias mostly toward the spring and summer seasons, with the winter period being

the most poorly sampled.

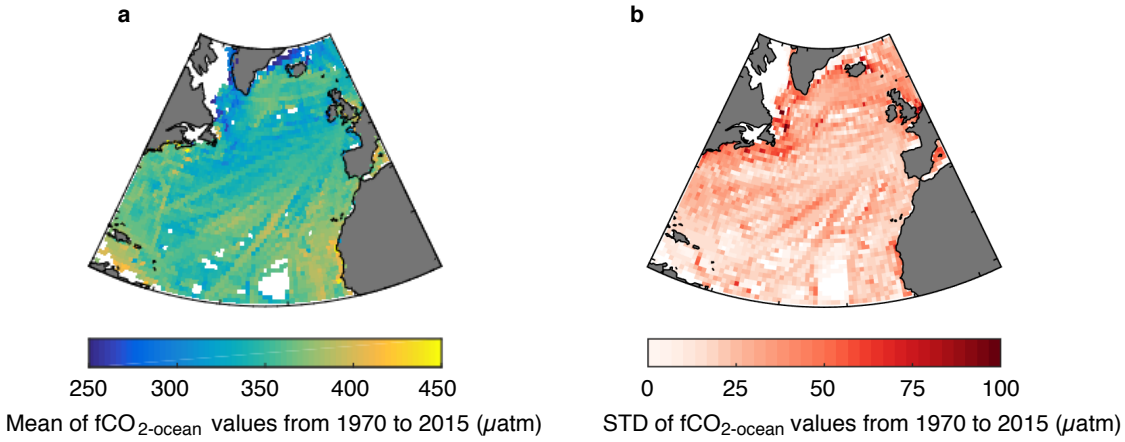

Figure 3.4: The North Atlantic surface $fCO_{2-ocean}$. Each grid cell represents the temporal average **a**, and standard deviation **b** of all $fCO_{2\text{-ocean}}$ value(s) available in that grid cell during the period 1970-2015 (based on SOCATv4).

In the North Atlantic, each SOCATv4 grid cell provides between 0 and 158 months with data spread across the period 1970-2015, with a median of 9 months. As such, while the most sampled grid cell (i.e. the one with 158 monthly $fCO_{2\text{-ocean}}$ values) covers ∼29% of the SOCATv4 period, 50% of the North Atlantic grid cells (i.e. the ones with 9 monthly $fCO_{2\text{-ocean}}$ values) only cover 1.6% of the available period. The relatively small amount of data in each of the North Atlantic grid cells most likely provide a noisy signal that prevents the detection of the oceanic response to increasing atmospheric CO_2 concentrations. Moreover, as each grid cell typically samples different months of different years, the temporally-averaged $fCO_{2\text{-ocean}}$ corresponds for each grid cell to a different "snapshot" in time, which leads to further challenges when understanding the mean $fCO_{2\text{-ocean}}$ patterns across the North Atlantic (Figure 3.4).

From the current observational coverage, the mean $fCO_{2\text{-ocean}}$ broadly presents higher values in the subtropical gyre than in the subpolar gyre, with particularly high values in the Canary upwelling system (Figure 3.4a). While the spatial $fCO_{2\text{-ocean}}$ patterns possibly describe different phases of internal variability due to the irregular temporal sampling (rather than describing a uniform mean state), they are coherent with discussions made in Chapter 1, which highlighted the impact of the North Atlantic circulation on the CO_2 system. For instance, by bringing enriched-DIC deep waters to the surface, upwelling areas are found to be regions

of CO_2 outgassing, which would correspond to regions of high $fCO_{2\text{-ocean}}$ relative to the CO_2 fugacity in the atmosphere (Figures 1.5; 3.4a).

3.1.2 Potential biases behind the $fCO_{2-ocean}$ observations

Unlike the previous chapter that provided extensive discussion of the various sources of uncertainty behind the GLODAP observational-based product (i.e. discussions which were particularly supported by personal sea-going experiences), the potential sources of uncertainty behind the SOCATv4 product are relatively limited for two main reasons: (1) $fCO_{2\text{-ocean}}$ is derived to a large degree from autonomous measurement systems (i.e. ship underway systems, fixed moorings, drifting buoys) (Bakker et al., 2016), which potentially limits the intrusion of experimental and judgement errors (as opposed to DIC or TA measurements which rely on manual sampling and measurements procedures; Appendix B), and (2) the $fCO_{2\text{-ocean}}$ values are subjected to limited transformation, for instance through the limited temporal average and the absence of spatial interpolation, which can lead otherwise to further challenges when interpreting and using the data (e.g. Section 2.1.1). Nevertheless, it is worth pointing out two potential sources of uncertainty that are linked to the experimental procedures of underway systems and may affect the resulting surface $fCO_{2\text{-ocean}}$.

Underway systems continuously pump seawater from the bottom of a ship (between 5 and 7 m depth; Sims et al., 2017) and circulate that seawater into an equilibrator on board, from which autonomous measurements are generated (Pierrot et al., 2009). Such protocols rely on the fact that (1) the depth at which seawater is collected does not precisely correspond to the surface, and (2) the temperature at which the measurement is made (i.e. the equilibrator temperature) does not equal the intake temperature. The impact of the sampling depth on the resulting $fCO_{2\text{-ocean}}$ was studied by Sims et al. (2017), in which they developed a near-surface ocean profiling buoy specifically designed to quantify the change in $fCO_{2\text{-ocean}}$ within the first 10 m (among other variables). Through two deployments, they identified a decreasing vertical gradient in $fCO_{2\text{-ocean}}$ of 4 µatm from 0.5 m to 5 m depth (mostly due to changes in temperature and salinity that affected the CO_2 solubility), corresponding to an overestimated air-sea CO_2 flux of $\sim$44% when using the 5 m $fCO_{2\text{-ocean}}$ value instead of 0.5 m value. While the *in-situ* experiments were limited to two sites and the resulting gradients did not persist through all hours of the day (Sims et al., 2017), near-profiling systems suggest that underway

seawater observations may lead to a substantial bias (i.e. underestimation of 4 µatm) in the deduced $fCO_{2\text{-ocean}}$ values.

To account for the temperature change between the collected and analysed seawater, a correction was applied in SOCATv4 (Takahashi et al., 1993), and when the intake temperature was not provided (on rare occasions), the "bulk" Optimum Interpolation Sea Surface Temperature (OISST) product (Reynolds et al., 2007) was taken instead. However, neither does the corrected sample's temperature or sub-surface temperature product characterise the ocean's skin layer (about 500 µm thick); introducing an error when SOCAT is used to calculate air-sea CO_2 flux (which occurs through the skin). Indeed, Robertson and Watson (1992) investigated the impact of the skin temperature effect (within a thin layer of about 500 µm), which is usually described by a cooling ($\sim$0.3°C) in the first millimetre of the surface layer compared to the rest of the mixed layer, on the air-sea CO_2 flux. They found that by considering the cooling at the surface skin, an additional air-sea CO_2 flux occurs, which was estimated between 0.3 and 1.3 GtC.yr^{-1}. As such, recent $fCO_{2\text{-ocean}}$ observational-based product (e.g. Schutler et al., 2016) use the skin temperature in their air-sea CO_2 calculations.

While the two experimental features in underway systems might counter-interact one another (the overestimation of the air-sea CO_2 flux from sampling the subsurface instead of the top layer might cancel the underestimation in the flux from the omission of the skin temperature correction), these examples highlight the sensitivity of the CO_2 surface system and the challenges that the observational community could tackle in the future.

3.1.3 Investigating the impact of sparse data on annual means and trends

While the extension of $fCO_{2\text{-ocean}}$ observations has contributed to improvement in quantifying and understanding the surface CO_2 system, the remaining gaps still hinder the complete understanding of the recent basin-wide North Atlantic change in the CO_2 sink. One may therefore ask: By how much does a non-ideal sampling scheme (either poor or relatively well-developed) impact annual means and trends? Evidently, the generation of annual means from a non-complete field will be different from the value which would have been returned from a perfectly sampled field, but the amplitude of this difference remains unclear. Without the existence of a spatially and temporally complete observational CO_2

system, such impact cannot be assessed in the real-world, but can however be investigated using the idealised conditions of a model-world, in which each grid cell of each month holds a $fCO_{2\text{-ocean}}$ value. Here, a model-based study provides a perspective on the impact of the SOCATv4 sampling scheme in the North Atlantic on basin-wide annual $fCO_{2\text{-ocean}}$ means, as well as on trends.

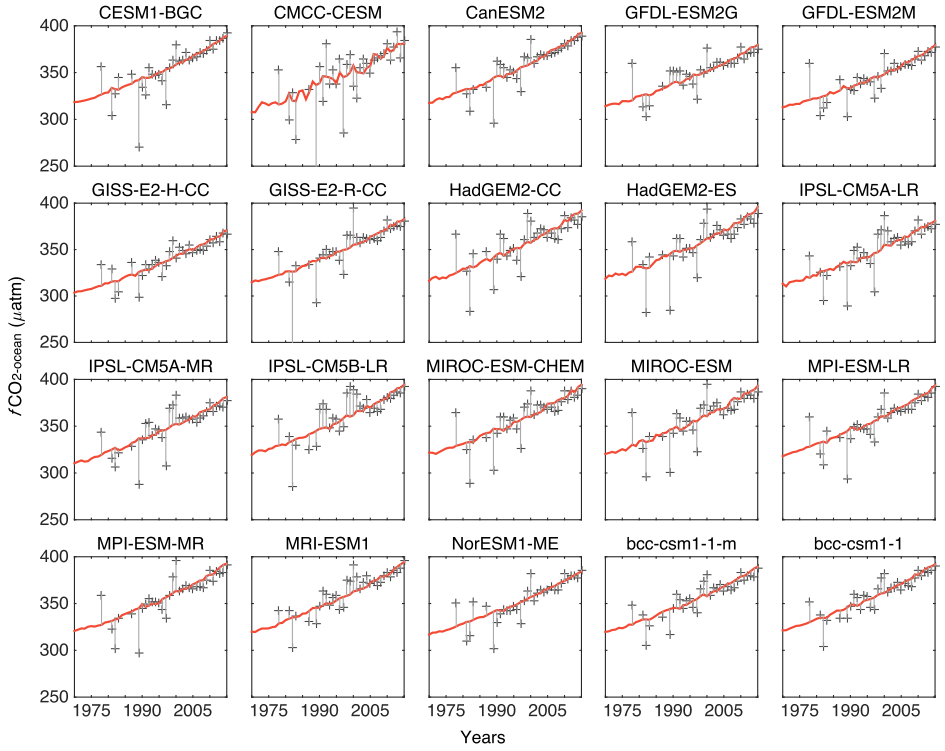

Figure 3.5: The North Atlantic $fCO_{2\text{-ocean}}$ sampling scheme's impact on annual means: a model-based study. Area-weighted North Atlantic annual $fCO_{2\text{-ocean}}$ means, calculated from the complete model field (red line) and from the subsampled field (crosses) within each CMIP5 model. The subsampled field refers to the complete model field that has been extracted at the same locations and times as provided by the SOCATv4 product, and therefore refers to the model version with the current observational coverage.

The present model-based study uses the North Atlantic $fCO_{2\text{-ocean}}$ field from 20 different CMIP5 models (Appendix A; Table A.3) and is based on two calculation steps that are repeated for each of the 20 models. First, the annual area-weighted means are calculated from 1970 to 2015 using the $fCO_{2\text{-ocean}}$ model field, which is hereinafter referred to as the "complete" field for emphazing reasons (red line in Figure 3.5). Second, the complete $fCO_{2\text{-ocean}}$ model field is sampled at the exact same locations and times at which the $fCO_{2\text{-ocean}}$ observations are available in SOCATv4, hereinafter referring to as the

"subsampled" field, and is then averaged into annual area-weighted means (crosses in Figure 3.5). For each model, the comparison between the annual means calculated from the complete and the subsampled $fCO_{2\text{-ocean}}$ fields allows visualisation of the impact of the relatively poor sampling in the 1970s or 1980s and of how the increase in number of measurements since the 2000s has improved our ability to capture the annual signal (Figures 3.5; 3.2b).

For the 46-year interval (i.e. 1970-2015) and across the 20 models, the error and absolute error between the annual means from the subsampled and complete fields are respectively -1.2 $\pm$ 1.6 µatm and 10.1 $\pm$ 2.4 µatm (where the standard deviation corresponds to the inter-model variability), which indicate that the sampling tends to underestimate the model truth. Such error values are however particularly impacted by the sampling in the early decades when observations were particularly sparse, and are reduced almost by half when the interval 2000-2015 is considered, for which the error and absolute error are -0.03 $\pm$ 1.2 µatm and 5.8 $\pm$ 1.5 µatm, respectively.

To study the impact of the observational coverage on our ability to quantify trends in time-varying surface ocean CO_2, $fCO_{2\text{-ocean}}$ linear trends are calculated using each subsampled (i.e. $\Gamma_{subsampled}$) and complete model fields (i.e. $\Gamma_{complete}$), and over intervals that vary from 2 to 46 years (i.e. 2013-2015 to 1970-2015, respectively) (Figure 3.6a). Each trend value is defined by a standard error, which captures the variability due to the sampling, the anthropogenic response to increasing atmosphere CO_2 and the internal biogeochemical change. Across the models, the standard errors on $\Gamma_{subsampled}$ are larger than on $\Gamma_{complete}$ (Figure 3.6b), demonstrating that it is not appropriate to simply quantify the temporal surface ocean CO_2 trends, even in the relatively well sampled North Atlantic, by simply averaging the available observations across space. Trends are also particularly sensitive to the sampling on short timescales (e.g. 2011-2015, 2012-2015), as the year-to-year differences in the sampling regime dominate the model-truth signal (Figure 3.6). The model-averaged standard errors on $\Gamma_{complete}$ also increases with the shortening of the interval of study (Figure 3.6b), indicating that substantial internal variability arises along those timescales.

Overall, about 68% of all $\Gamma_{subsampled}$ are not distinctive from $\Gamma_{complete}$ (i.e. trends with

associated standard error overlap; non-dots in Figure 3.6a). While such results would indicate that the current CO_2 observational system captures most of the model-true signal, the standard errors associated with all $\Gamma_{subsampled}$ are larger than for all $\Gamma_{complete}$ (between 11 and 30% larger; Figure 3.6b). For instance, the $\Gamma_{subsampled}$ for the CESM1-BGC model and the interval 1992-2015 is 2.02 $\pm$ 0.30 µatm·yr^{-1} and the corresponding $\Gamma_{complete}$ is 1.93 $\pm$ 0.03 µatm·yr^{-1}. Since the atmospheric fCO_2 trend over a similar interval (1992-2014) is 1.88 $\pm$ 0.02 µatm·yr^{-1} (Chapter 5), the large standard error on $\Gamma_{subsampled}$ (i.e. 0.30 µatm·yr^{-1}) indicates that it would remain impossible to determine from the current observing system whether the North Atlantic CO_2 sink, over that interval and for that model, increased or decreased.

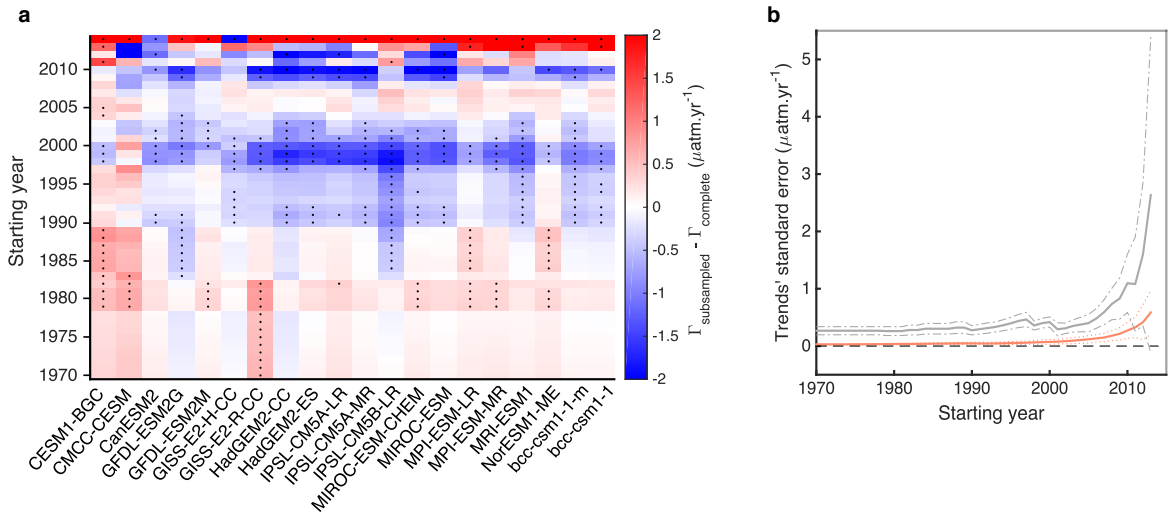

Figure 3.6: The North Atlantic $fCO_{2\text{-ocean}}$ sampling scheme's impact on trend: a model-based study. **a**, Difference between the $fCO_{2\text{-ocean}}$ linear trend calculated from the subsampled field ($\Gamma_{subsampled}$) and the one calculated from the complete field ($\Gamma_{complete}$), where a negative difference (blue) thus indicates that the sampling coverage tends to underestimate the true model trend. Dots indicate when the two trends and their associated standard errors do not overlap, indicating that their difference is significant at 1σ. **b**, Model-average (thick) of the standard errors of the $fCO_{2\text{-ocean}}$ trends calculated from the subsampled (grey) and complete (orange) fields for each interval configuration (x-axis), with the models $\pm 1\sigma$ from the model-average (dashed). The starting year in **a** and **b** refers as the first year of the period over which the trends were calculated, with the 2015 being the ending year.

Another study (Tjiputra et al., 2014) took advantage of the perfectly known models to assess the impact of the observational CO_2 coverage on $pCO_{2\text{-ocean}}$ trends across the period 1970-2011 and over 14 regions of the globe (specifically using SOCATv2; Bakker et al., 2014). Tjiputra et al. (2014) identified that, in all regions, the model-mean trend from the

subsampled field was lower than the model-mean trend from the complete field, which is the opposite result from the current illustrative analysis (i.e. here, the model-mean trend for the period 1970-2015 in the North Atlantic is higher from the subsampled than the complete fields[3]). These diverging results are explained by the different methods used between the studies to generate the trends:

- Here: For each month of the period 1970-2015, the subsampled $fCO_{2\text{-ocean}}$ monthly field is averaged (area-weighted), from which annual means are deduced and from which linear trends are fitted. As such, this method is particularly sensitive to the month(s) at which observations were recorded within a year.

- Tjiputra et al. (2014): For each grid cell, $pCO_{2\text{-ocean}}$ anomalies are generated by removing the cell's mean value calculated over 1970-2011. For each month of that period, the subsampled $pCO_{2\text{-ocean}}$ anomalies are averaged (area-weighted). Based on that time series, the seasonal cycle is calculated and removed from the time-varying $pCO_{2\text{-ocean}}$, which is finally used for the linear fit. While such approach allows to remove some aliasing due to the sampling, it also substantially relies on the sampling capturing the climatological and seasonal states, which (as discussed earlier) is not assured and might therefore add another source of uncertainty.

Whether the subsampled trend in the models underestimates or overestimates the model true trend, these examples clearly show the difficulties in capturing the amplitude of the true trend from the current observational coverage. However, while models provide a useful platform to explore uncertainties that would remain impossible to quantify (to this day) in the real-world, applying the identified uncertainties and potentially correcting the observational-based statistical results accordingly, would rely on models capturing the spatial and temporal variabilities of the real word biogeochemistry, and in particular of the surface $fCO_{2\text{-ocean}}$.

3.2 Model evaluation

By gathering surface $fCO_{2\text{-ocean}}$ observations made over the past four decades into monthly gridded means and thus by preventing the intrusion of source of errors from interpolation

[3]Specifically, the model-mean trends from the subsampled and complete fields are 1.70 ± 0.17 µatm·yr^{-1} and 1.56 ± 0.06 µatm·yr^{-1}, respectively, with the $\pm 1\sigma$ representing the inter-model variations.

methods (Bakker et al., 2016), the monthly gridded SOCATv4 product provides data in a format that is valuable for model evaluation, but lends itself to do point-by-point evaluation, which (1) could be an overly harsh comparison with models if the models were to not simulate the real-world spatial features, and (2) does not allow direct evaluation of the models' characteristics in which the community is particularly interested, such as the simulated basin-wide CO_2 ocean uptake. This section aims to evaluate the North Atlantic surface $fCO_{2\text{-ocean}}$ field of 20 CMIP5 models (Appendix A.1; Table A.3) against the monthly gridded SOCATv4 product (referred as SOCATv4; Bakker et al., 2014), and particularly investigating the models' performance at different temporal variability (overall mean, seasonal and interannual variability; Section 3.2.2). To highlight the importance of subsampling the models' $fCO_{2\text{-ocean}}$ fields for model-observation comparisons, the model evaluation's methods of two studies (Tjiputra et al., 2014 and Arruda et al., 2015) are first discussed (Section 3.2.1).

3.2.1 The importance of subsampling the models: previous studies

To provide coherent comparisons between modelled and observational fields (e.g. Section 3.1.3), models should ideally be subsampled at the same locations and times as the observations. Nevertheless, some model evaluation studies compared $fCO_{2\text{-ocean}}$ means from the sparse observational coverage with means from the complete model fields, which could impact the assessment of the model's performance (depending on the period and/or region of study). For instance, Tjiputra et al. (2014) evaluated the spatial distribution of the global mean $pCO_{2\text{-ocean}}$ of five CMIP5 models with the monthly gridded SOCAT version 2 (referred as SOCATv2; Bakker et al., 2014) over the period 1991-2011, without any preliminary subsampling of the $pCO_{2\text{-ocean}}$ model fields. While the chosen period (i.e 1991-2011) covers most of the available data within SOCATv2 (Bakker et al., 2014), model evaluation based on the model fields without prior subsampling would likely lead to substantially different model skill results, especially in regions with limited available observations, such as the Southern Ocean. Indeed, while Tjiputra et al. (2014) identified that the models poorly captured the magnitude of the surface $pCO_{2\text{-ocean}}$ (e.g. overestimation in the subtropical North Atlantic and parts of the Southern Ocean), a somewhat similar analysis shows that when subsampling the model field following the observational coverage (e.g. from the MPI-ESM-LR model; Figure 3.7), the means from the complete model field overestimates the mean from the

subsampled model field in the majority of the Southern Ocean; suggesting (with further analysis) that this model actually underestimates the observed mean values in that basin (which is the opposite result from what was found in Tjiputra et al. (2014) study).

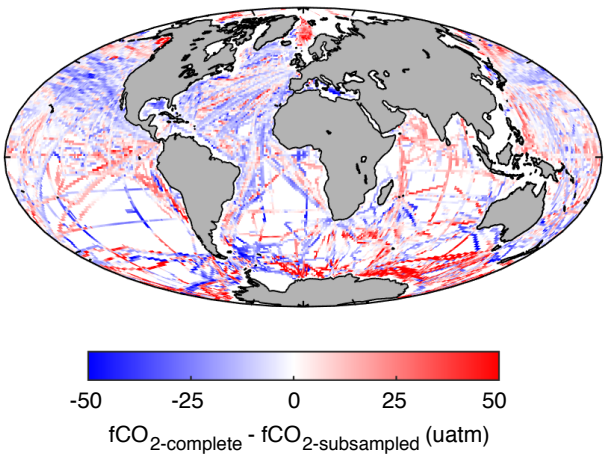

Figure 3.7: Approaches when spatially evaluating a model's $fCO_{2\text{-ocean}}$ field. Difference in the $fCO_{2\text{-ocean}}$ mean calculated over the period 1991-2011 between the mean calculated using the complete model field from the MPI-ESM-LR model and the mean calculated after subsampling the MPI-ESM-LR model field at the same locations and times as the SOCATv4 monthly gridded product (note that Tjiputra et al. (2014) used the SOCATv2 monthly gridded product; Bakker et al. (2014)). The MPI-ESM-LR model was one of the five models used in Tjiputra et al. (2014).

Due to the irregular seasonal distributions in the $fCO_{2\text{-ocean}}$ observations (e.g. in the North Atlantic in Figure 3.3), the preliminary model-subsampling step is also crucial when evaluating the models' $fCO_{2\text{-ocean}}$ seasonality; a point that is discussed with the model evaluation carried in Arruda et al. (2015). Arruda et al. (2015) specifically evaluated (without preliminary subsampling) the $pCO_{2\text{-ocean}}$ seasonal variability from a regional ocean-biogeochemical model, which was initialised and forced with climatological physical and biogeochemical values and set in Argentinian coastal waters, against the seasonality captured by the SOCATv2 product in that area (Bakker et al., 2014). By doing so, the authors assumed that the simulated and observed seasonal $pCO_{2\text{-ocean}}$ climatologies describe similar periods, which is not evident as the period over which the climatological forcing fields are deduced from (which should correspond to the simulated climatological $pCO_{2\text{-ocean}}$ field) is not indicated. However, even if the model were to simulate the $pCO_{2\text{-ocean}}$ climatology for the period 1970-2011 (i.e. the SOCATv2 period), it would still relies on the SOCATv2 coverage in the Argentinian coastal waters being well-enough distributed (both

spatially and temporally) to capture the seasonal climatological state. Such assumption is questionable, especially since the SOCATv2 observational coverage in the variable Argentinian coastal waters provides 30% of the summer months (JFM) with at least one $pCO_{2\text{-ocean}}$ value and 18% for the winter months (JJA). Furthermore, as the model simulation was run at a constant atmospheric CO_2, the comparison between the simulated oceanic response to this constant atmospheric CO_2 and the real-world ocean which does respond to a changing atmospheric CO_2 may not be valid. Despite these points, Arruda et al. (2015) suggested that their model simulation captured reasonably well the seasonality of the surface $pCO_{2\text{-ocean}}$.

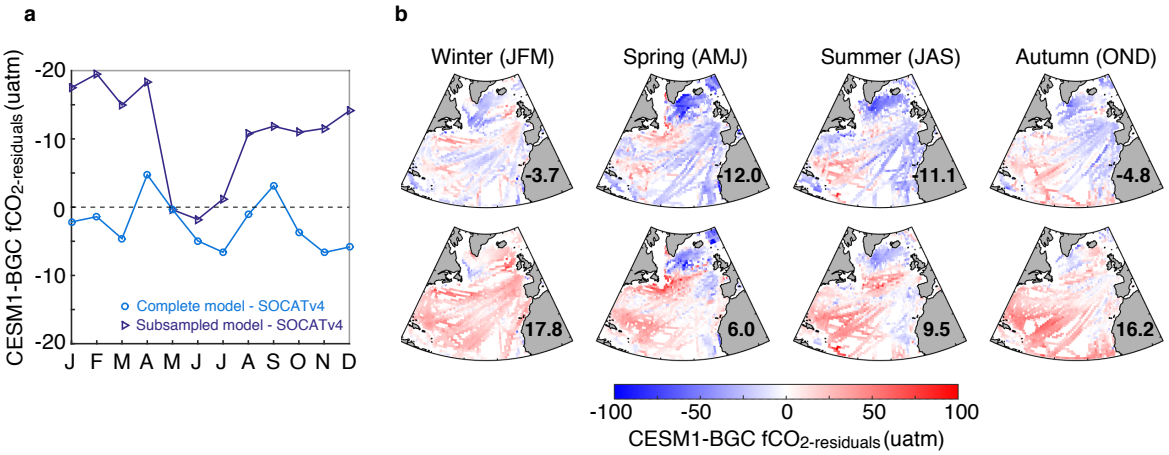

Figure 3.8: Approaches when evaluating a model's $fCO_{2-ocean}$ seasonal signal: example in the North Atlantic. **a**, Difference between the seasonal cycle calculated from the complete/subsampled $fCO_{2\text{-ocean}}$ model field and the seasonal cycle calculated from SOCATv4 (in light/dark blue, respectively). **b**, Spatial differences between mean seasonal $fCO_{2\text{-ocean}}$ from the the complete/subsampled field and the mean seasonal $fCO_{2\text{-ocean}}$ from SOCATv4 (in top/bottom maps, respectively). The number in each figure corresponds to the mean error (area-weighted) of each residual map (µatm). The randomly chosen model is CESM1-BGC and all climatologies have been calculated using data from 1970-2015.

To highlight the impact of the subsampling processing step on a seasonal model evaluation study, the North Atlantic seasonal $fCO_{2\text{-ocean}}$ from the complete and from the subsampled CESM1-BGC model field (this model was randomly chosen among the other 19 CMIP5 models) are compared to the North Atlantic seasonal $fCO_{2\text{-ocean}}$ from SOCATv4 (Figure 3.8). Specifically, the seasonal cycle (Figure 3.8a) and the seasonal spatial patterns (Figure 3.8b) of the $fCO_{2\text{-residuals}}$ are explored (i.e. $fCO_{2\text{-residuals}} = fCO_{2\text{-model}} - fCO_{2\text{-SOCATv4}}$, where $fCO_{2\text{-model}}$ is either the complete or subsampled model field). A model evaluation based on the complete

model field would suggest that the CESM1-BGC model captures the seasonal cycle relatively well (the overall mean error on the seasonal cycle, which is the average of the difference between the modelled and the observed seasonal cycles, is -2.5 $\pm$ 3.7 µatm; Figure 3.8a), but with a tendency to underestimate the observed values especially in the subpolar region and the eastern North Atlantic (negative mean errors in Figure 3.8b). On the other hand, a model evaluation based on the subsampled model field would qualify the model with a rather poor skill (the mean error on the seasonal cycle is 10.7 $\pm$ 7.3 µatm; Figure 3.8a), especially in winter and autumn, where the model overestimates the observed $fCO_{2\text{-ocean}}$ value across most of the basin (Figure 3.8b). As such, depending on whether the complete or subsampled model field is used for the evaluation analysis, substantial contradicting model assessments can rise, suggesting that the model evaluation carried in Arruda et al. (2015) would potentially return different results if the model output was initially subsampled.

These two illustrative examples highlight the importance of comparing mean properties deduced from the same spatial and temporal coverage as in the real-world and therefore provides context for the CMIP5 model-evaluation analysis that follows.

3.2.2 Evaluating the North Atlantic surface $fCO_{2-ocean}$ in the CMIP5 models

Here, the North Atlantic surface $fCO_{2\text{-ocean}}$ of 20 CMIP5 models (Appendix A.1; Table A.3) are evaluated against the SOCATv4 monthly gridded product (Table A.5), focusing on the spatial mean patterns and on the annual and seasonal variabilities. For consistency reasons, each CMIP5 model has been subsampled at the same locations and times as the observations gathered in SOCATv4. Throughout the rest of this section, the $fCO_{2\text{-residuals}}$ will refer, at each grid cell over the period 1970-2015, to the surface $fCO_{2\text{-ocean}}$ difference between the subsampled model and SOCATv4; where a positive $fCO_{2\text{-residuals}}$ would indicate that the model overestimates the observed value. For simplicity reasons, the error on the measurements is assumed to be minimal and is therefore excluded from the subsequent analysis.

Taylor diagram

As discussed in the previous chapters (Sections 1.5 and 2.2.2), Taylor diagrams are powerful compact visuals that allow one to easily differentiate skilled models from the poorly skilled

ones, through the quantitative evaluation of their variance, correlation and unbiased RMSE with the observational field. The surface $fCO_{2\text{-ocean}}$ Taylor diagram shows that the CMIP5 models overall present a similar behaviour to each other, except for the CMCC-CESM model which clearly stands out from the main cluster (Figure 3.9). Indeed, the CMCC-CESM has an unbiased RMSE of 73.2 µatm and a standard deviation of 78.0 µatm, while the other models have on average the respective values of 35.2 $\pm$ 7.1 µatm and 37.7 $\pm$ 5.5 µatm, where the $\pm 1\sigma$ represents the inter-model variability. Considering that the observed $fCO_{2\text{-ocean}}$ variance is 28.5 µatm, all models overestimate the observed variance, from 2.2 µatm with the GFDL-ESM2M model to 49.5 µatm with CMCC-CESM. The CMIP5 models also present a weak to poor positive correlation with SOCATv4, from $\sim$0.70 with the CESM1-BGC and GFDL-EMS2M models to 0.18 with the GISS-E2-H-CC model. By also presenting the smallest unbiased RMSE (22.8 µatm) of all the CMIP5 models, GFDL-ESM2M appears through the Taylor diagram analysis to be the model with the highest $fCO_{2\text{-ocean}}$ skill. However, since Taylor diagrams do not provide information on the potential model bias, further analysis should be done to fully explore the models' $fCO_{2\text{-ocean}}$ performance.

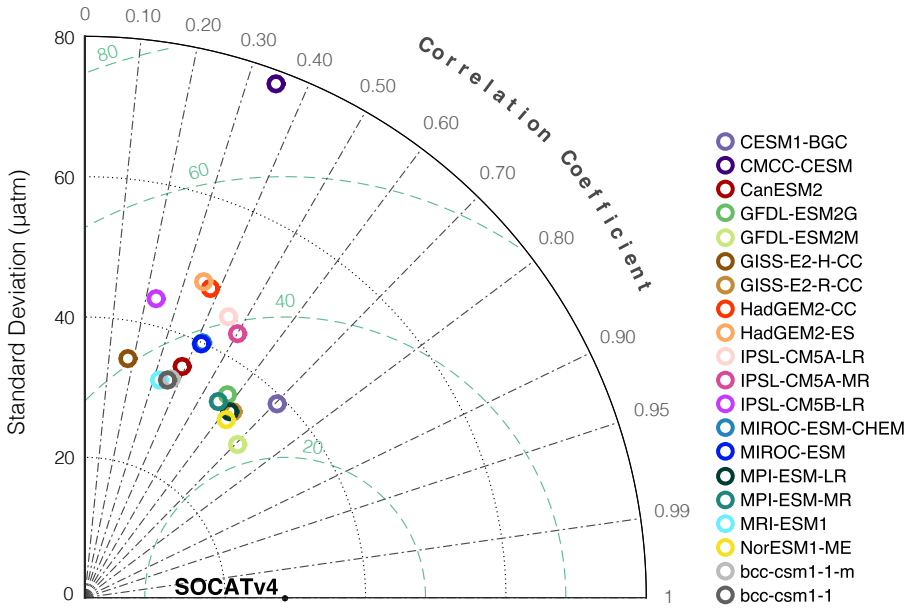

Figure 3.9: Surface $fCO_{2\text{-ocean}}$ **Taylor diagram in the North Atlantic**. Standard deviation (radial axis), correlation (polar angle) and unbiased RMSE (green lines) between each subsampled CMIP5 model and SOCATv4, using all data available during the period 1970-2015 (Taylor, 2001).

Surface maps

To explore the models' spatial performance, the temporally-averaged $fCO_{2\text{-residuals}}$ are studied (Figure 3.10). On average, the CMIP5 models overestimate the observed $fCO_{2\text{-ocean}}$ values in the subtropical region and in most of the subpolar region, except around the Irminger and Labrador Seas where the models tend to underestimate the observations. A localised enhanced overestimation is found in most models around the Gulf Stream area (e.g. CanESM2, MIROC-ESM-CHEM), suggesting that these models might provide a shift in the location of their gyres boundaries, compared to the real-world mean boundaries. The amplitude of the $fCO_{2\text{-residuals}}$ is quite variable across the models, with relatively low amplitudes for the GFDL-ESM2M model and high for the CMCC-CESM model, which are characterised with a relatively low (17.1 µatm) and high (46.3 µatm) Mean Absolute Error (MAE), respectively (Figure 3.10).

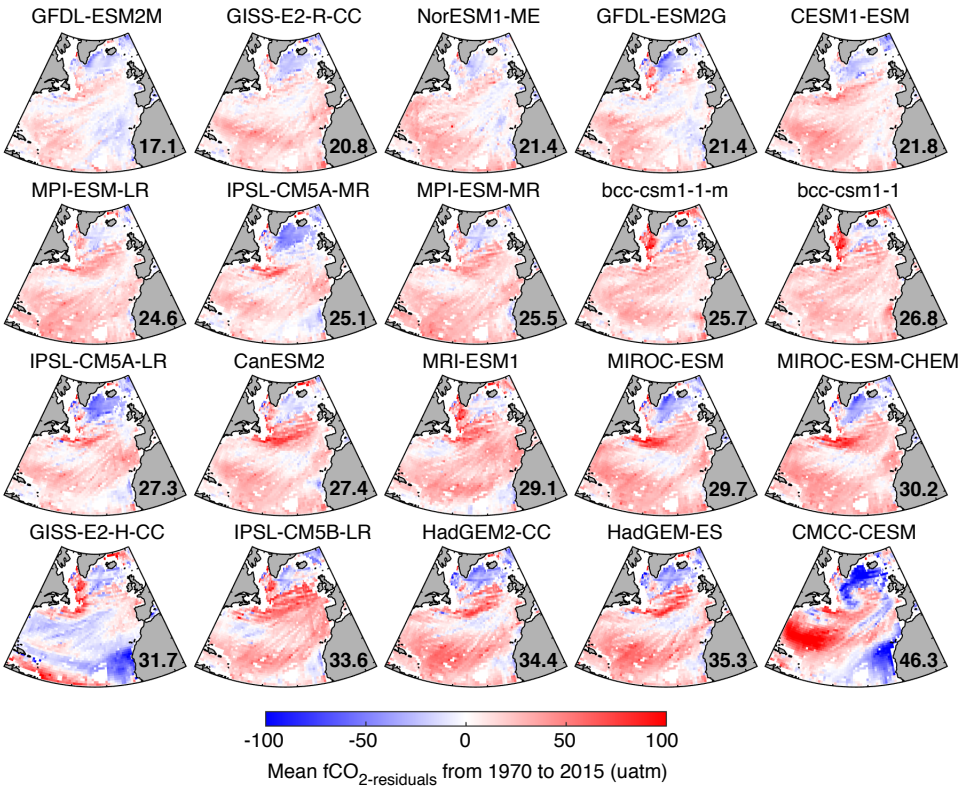

Figure 3.10: Mean surface $fCO_{2\text{-residuals}}$ in the North Atlantic. For each CMIP5 model, average of the $fCO_{2\text{-residuals}}$ calculated over the period 1970-2015 from subsampled data, where red shows where the model overestimate, on average, the SOCATv4 values. The number indicated in each map corresponds to the area-weighted MAE of the $fCO_{2\text{-residuals}}$ values displayed in each map (in µatm). The models are ordered in increasing MAE from the top left corner to the bottom right.

To constrain the average $fCO_{2\text{-residuals}}$ spatial patterns (Figure 3.10), the standard deviations around the spatial means are presented (Figure 3.11). Indeed, even if a model provides a mean $fCO_{2\text{-residuals}}$ that is close to zero, which would indicate that the model captures the mean observed $fCO_{2\text{-ocean}}$ well (e.g. in a grid cell, two $fCO_{2\text{-residuals}}$ values of -100 µatm and 100 µatm would return an ideal mean of 0 µatm), a relatively high standard deviation in the $fCO_{2\text{-residuals}}$ (e.g. 141 µatm from the above example) would suggest that the model poorly captures the observed variability.

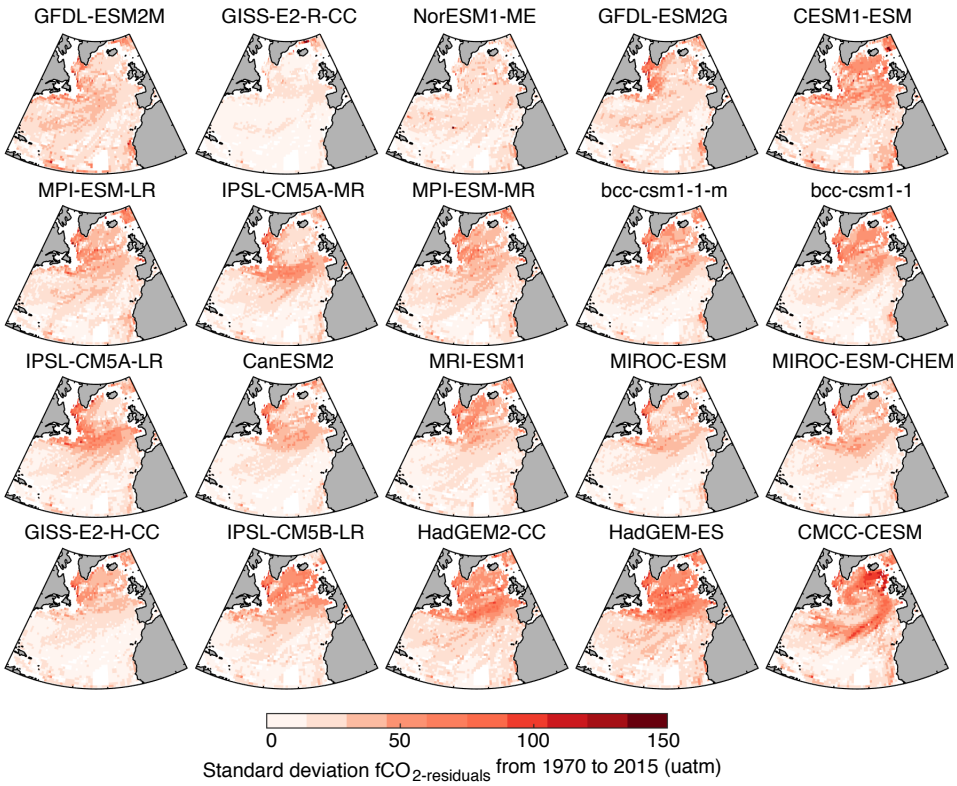

Figure 3.11: Standard deviation of the mean surface $fCO_{2\text{-residuals}}$ in the North Atlantic. For each CMIP5 model, standard deviation from the average $fCO_{2\text{-residuals}}$ calculated over the period 1970-2015 from subsampled data. The models are displayed in the same order as in Figure 3.10.

Most of the CMIP5 models present lower standard deviations in the subtropical region than in the subpolar basin (e.g. HadGEM2-CC, IPSL-CM5A-LR; Figure 3.11), which highlights that (1) the models' overestimation in the subtropical gyre (Figure 3.10) is likely to be maintained throughout the observations recorded within a subtropical grid cell, and (2) these models overall fail at capturing the subpolar $fCO_{2\text{-ocean}}$ variability, at least for the times at which the observations were made. The disparities between the subpolar $fCO_{2\text{-ocean}}$ in the

concerned models and in the observations could be explained by the fact that the observations were made at times of great variability, which the models fail to capture, and/or by poorly simulated key physical/biogeochemical mechanisms in the models, such as the $fCO_{2\text{-ocean}}$ seasonal cycle which will be investigated later in the section. Nevertheless, the GFDL-ESM2M, GISS-E2-R-CC and NorESM1-ME models present a rather uniform standard deviation spatial pattern across the North Atlantic, with relatively low values (median value of 18 µatm), indicating that those models not only capture the mean spatial $fCO_{2\text{-ocean}}$ distribution well, but also provide characteristics in the subpolar gyre that enable the representation of the real-world $fCO_{2\text{-ocean}}$ variability captured by the observational coverage.

Annual means

The North Atlantic $fCO_{2\text{-ocean}}$ temporal behaviour in the CMIP5 models is first examined through annual averages. Specifically for each model, the monthly $fCO_{2\text{-residuals}}$ available for the period 1970-2015 are first spatially averaged (using area-weighted means), which are then averaged into annual values (Figure 3.12). Across the models, the amplitude of the annual $fCO_{2\text{-residuals}}$ is maximal for the low sampled years (e.g. the early 1980s or 1999-2000; Figure 3.12), which indicates that the models do not perform well on localised observations. For the highly sampled years (e.g. since 2003), the CMIP5 models tend to consistently overestimate the SOCATv4 values, leading to an overall positive mean error throughout the period 1970-2015; except for the GISS-E2-H-CC model which provides an overall negative mean error (as indicated by the top numbers in Figure 3.12). While in the above analyses the CMCC-CESM model was identified with a poor $fCO_{2\text{-ocean}}$ skill, here the mean error through the period 1970-2015 for this model is the second lowest mean error in amplitude relative to the other models (Figure 3.12). However when analysing the year-to-year $fCO_{2\text{-residuals}}$, the CMCC-CESM model presents the highest annual mean errors in amplitude (from -53 to 39 µatm) compared to the other CMIP5 models, indicating the overall mean error only translates the cancellation of substantial biases (third column from the left in Figure 3.12). On the other hand, the GFDL-ESM2M presents, in amplitude, both small annual $fCO_{2\text{-residuals}}$ means throughout the period 1970-2015 and the smallest overall mean error (i.e. 2.1 µatm; second column from the left in Figure 3.12).

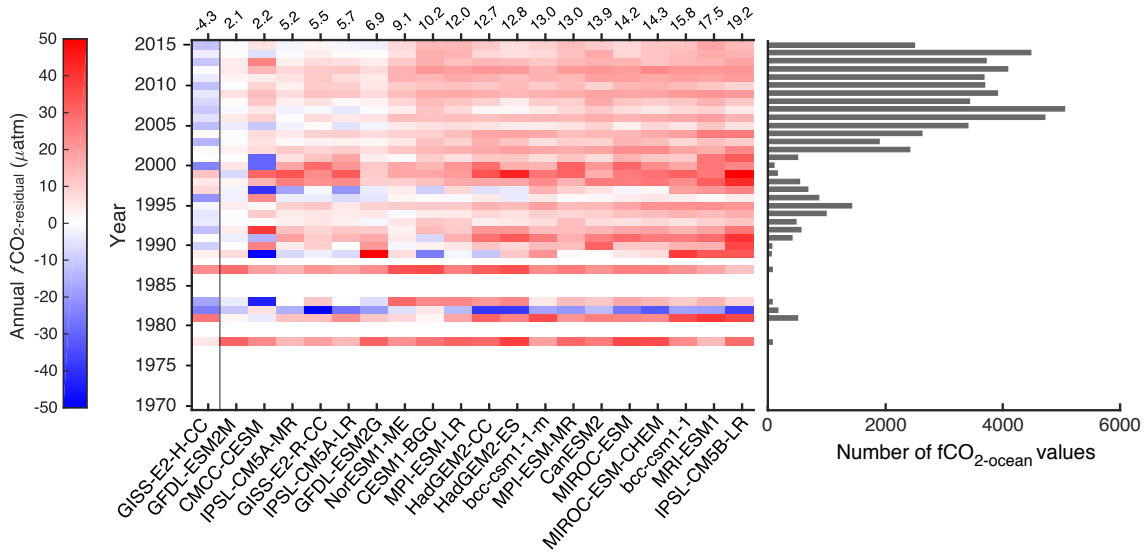

Figure 3.12: $fCO_{2\text{-residuals}}$ annual means in the North Atlantic. For each CMIP5 model, the annual area-weighted $fCO_{2\text{-residuals}}$ mean is displayed in colour, where red indicates that on average the model overestimates the observations during the year indicated in the y-axis. Models are displayed on the x-axis from the model with the lowest mean error to the model with the highest mean error (i.e. for each model the mean error, which is in µatm, is deduced by averaging all annual $fCO_{2\text{-residuals}}$ values within the period 1970-2015 and is indicated on top of the each vertical block). On the right-hand side, the bar plot recalls the number of SOCATv4 $fCO_{2\text{-ocean}}$ values per year, in the North Atlantic.

Seasonal cycle

To study the North Atlantic seasonal cycle of the surface $fCO_{2\text{-ocean}}$ without the interference of the long-term response due to the increasing atmospheric CO_2 concentration, each $fCO_{2\text{-ocean}}$ value should be initially detrended. However, the spatio-temporal gaps in the SOCATv4 product prevent us from knowing the long-term trend in $fCO_{2\text{-ocean}}$ at each grid point (as illustrated in Section 3.1.3) and therefore means it is not obvious how to detrend the data. To circumvent this issue, the atmospheric CO_2 fugacity (i.e. $fCO_{2\text{-atmosphere}}$), which is known at each location and output frequency in the CMIP5 models and in the observation-based GLOBALVIEW-CO_2 reference matrix (a product that interpolates atmospheric CO_2 measurements collected on specific sites across the globe) (GLOBALVIEW-CO2, 2013), is used as a proxy for the long-term change in $fCO_{2\text{-ocean}}$. The detrended seasonal cycle for each (subsampled) CMIP5 model and for the SOCATv4 data were obtained as follows:

- Determining the $fCO_{2\text{-atmosphere}}$. The $fCO_{2\text{-atmosphere}}$ was calculated from the atmospheric CO_2 mixing ratio (xCO_2), which was converted into $pCO_{2\text{-atmosphere}}$ and then into $fCO_{2\text{-atmosphere}}$ using the Equations 1.2 and 1.1, respectively. The

model-specific and observed xCO_2, sea surface temperature and sea level pressure fields required in those equations are described in Appendices A.1 and A.2, respectively. However, some of the observational products were not available for the entire SOCATv4 period (e.g. while the SOCATv4 monthly gridded product started in 1970, the Sea Surface Temperature (SST) product was available from 1982; Reynolds et al., 2007). As such, the $fCO_{2\text{-atmosphere}}$ was calculated for the period 1982-2015 for the observations and for the period 1970-2015 for the models (both available at a monthly frequency).

- Removing the long-term trend. For the observation-based and for each modelled $fCO_{2\text{-atmosphere}}$ products, a linear fit was applied to the $fCO_{2\text{-atmosphere}}$ time series contained within each North Atlantic grid cell. Each linear fit was then removed from the $fCO_{2\text{-ocean}}$ time series of the corresponding cell, generating the "detrended $fCO_{2\text{-ocean}}$". Since the linear fit in the observed-based $fCO_{2\text{-atmosphere}}$ is defined over 1982-2015 and SOCATv4 over 1970-2015, the $fCO_{2\text{-atmosphere}}$ linear relationship (i.e. trend and intercept) was used to extrapolate the time-varying $fCO_{2\text{-atmosphere}}$ prior 1982.

- Calculating the seasonal cycle. For each month of the period 1970-2015, the detrended North Atlantic $fCO_{2\text{-ocean}}$ was averaged into area-weighted means. The seasonal cycle was finally calculated by averaging the detrended $fCO_{2\text{-ocean}}$ means from the same month within each year (with data) over the period 1970-2015.

Based on the current observational coverage in the North Atlantic, the observed $fCO_{2\text{-ocean}}$ seasonal cycle (Figure 3.13) presents 3 main phases: (1) an overall decrease of about 26 µatm from October to May, containing a slight increase of about 1 µatm from January to March which could be linked to the sampling scheme, (2) an intense increase of about 26 µatm from May to August, and (3) a stagnation from August to October (Figure 3.13). Seasonal changes in the $fCO_{2\text{-ocean}}$ are likely to be linked to seasonal changes in surface temperature, vertical motion and biological activity (Mongwe et al., 2016). For instance, during the transition from winter to spring, the warming of the surface waters stimulates the biological activity and shoals the mixed-layer, which by concentrating nutrients-enriched waters toward the euphotic zone also enhances the biological activity. As such, spring blooms actively consume CO_2 through photosynthesis and therefore lead to a

decrease in surface $fCO_{2\text{-ocean}}$ (Chapter 1). However from spring to summer, the on-going warming of the surface waters tends to dominate the $fCO_{2\text{-ocean}}$ signal[4], by decreasing the CO_2 solubility in seawater and by therefore increasing the surface $fCO_{2\text{-ocean}}$ (Chapter 1). To capture the North Atlantic $fCO_{2\text{-ocean}}$ seasonal variability, the CMIP5 models should therefore simulate the mechanisms behind the biological and physical-dominant seasonal phases, as well as the timing of the transition's phases.

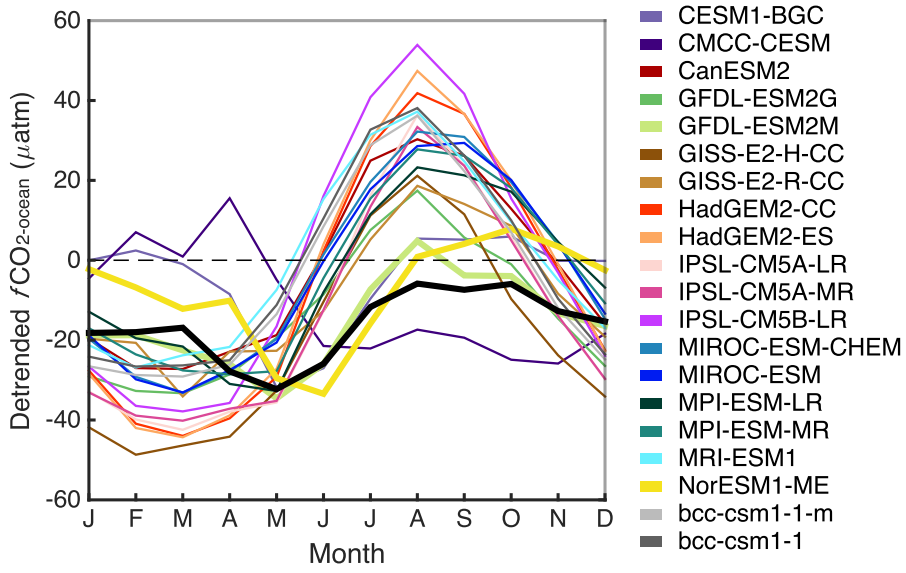

Figure 3.13: The $fCO_{2\text{-ocean}}$ seasonal cycle in the North Atlantic. Detrended seasonal cycle (see text for details) for the SOCATv4 (thick black) and for each CMIP5 models (colours). To make the GFDL-ESM2M and NorESM1-ME more distinguishable from the others models, their seasonal cycle is displayed in thick coloured lines.

From the visual comparison between the CMIP5 and the observed seasonal cycles (Figure 3.13), two model groups seems to appear: a first group made of the GFDL-ESM2M and the NorESM1-ME models (highlighted by thicker coloured lines on Figure 3.13), which are in good agreement with the phase and amplitude of the observed seasonal cycle (particularly the GFDL-ESM2M model with a Mean Absolute Error (MAE) of 2.8 µatm against 10.4 µatm for the Norwegian model; Figure 3.17); and a second group made of all the other CMIP5 models, which overall overestimate the amplitude of the observed seasonal variability (the model-mean MAE is 17.7 $\pm$ 4.1 µatm). The second model group provides a cycle that seems to follow a temperature-driven $fCO_{2\text{-ocean}}$ seasonal curve, which smoothly

[4]In certain areas (e.g. northeast Atlantic) and potentially under certain conditions, the biological effect on the air-sea CO_2 flux can be maximal in summer, as shown in Kortzinger et al. (2008).

transitions from minimal values in February-March (high CO_2 solubility in cold seawater) to peaks in August (low CO_2 solubility in warm seawater). To improve our understanding of the seasonal modelled behaviour in the North Atlantic, further seasonal analysis is investigated within the subtropical and subpolar regions separately.

The subtropical and subpolar regions are characterised by different SST regimes, with overall warm waters in the subtropics and cold waters in the subpolar. The transition between the two regions is therefore identified by a maximum in the latitudinal SST gradient, which broadly represents the location of the NAC (Chapter 1). To study the subtropical and subpolar $fCO_{2\text{-ocean}}$ seasonal variability with limited contamination from the shifts in the NAC and from the diverse representations of the NAC by the CMIP5 models, a zone of transition including most of the variability of each of the 20 CMIP5 models and of the real-world was determined using the following steps:

1. The climatology in the SST was calculated over the period 1982-2015 separately for each CMIP5 model and for the observation-based OISST product. The interval 1982-2015 was chosen for consistency reasons as it corresponds to the period of the observational SST product (Reynolds et al., 2007) (Appendix A.1 and A.2).

2. To identify the maximum change in the mean SST between the North Atlantic subtropical and subpolar regimes, the latitudinal gradient between each grid cell of each climatological SST field (i.e. the climatological field for each CMIP5 model and for the real-world) was calculated. Specifically, for a grid cell of longitude *i* and latitude *j*, the temperature gradient at the grid cell (*i*, *j*) was calculated by dividing the temperature difference between the values stored in cells *j+1* and *j-1* by the distance between the cells *j+1* and *j-1*, using an Earth radius of 6,400 km. A grid cell at latitude *j* with a negative latitudinal SST gradient (blue areas in Figure 3.14) therefore indicates that the temperature in the grid cell at latitude *j+1* is colder than the temperature in the grid cell at latitude *j-1*, at the same longitude *i*. As such, areas with the most negative latitudinal SST gradient correspond to areas of intense northward cooling, potentially corresponding to the shift between the subtropical to the subpolar gyres. To circumvent the identification of a strong northward cooling area beyond the vicinity of the NAC at the gyres boundaries (e.g. in the East Greenland Current, top plot in Figure

3.14), the maximum in the latitudinal SST gradient was therefore investigated within the region 25°N and 55°N of the North Atlantic (red dashed lines in Figure 3.14). The transition was therefore determined for the real-world and for each CMIP5 model as the most negative latitudinal temperature gradient within 25°N and 55°N for each longitude (black dots in Figure 3.14).

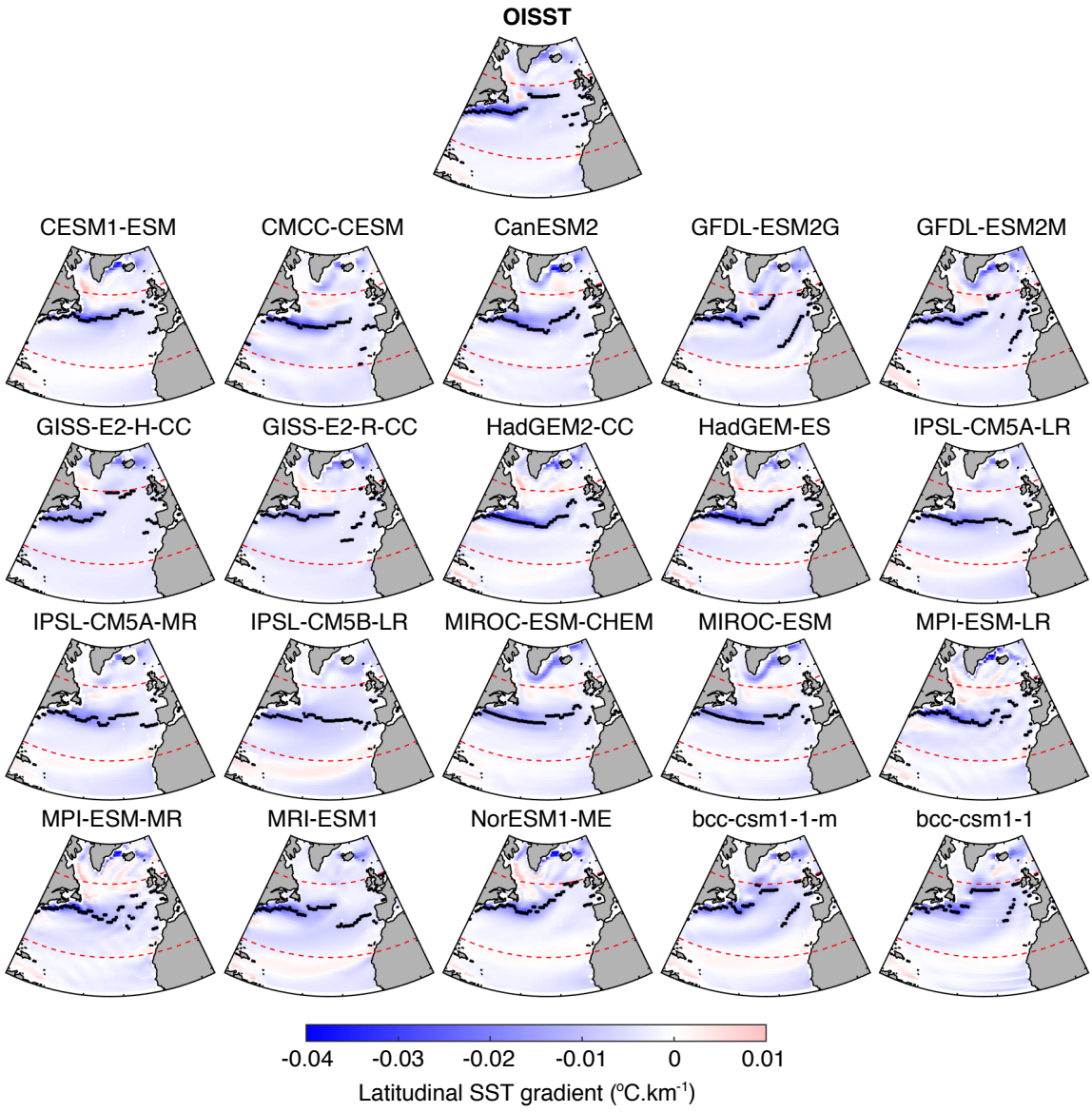

Figure 3.14: Northward change in the mean SST. Latitudinal SST gradient calculated as followed by step 2 in list above, based on the SST climatology calculated over the period 1982-2015 using the OISST observation-based product (Appendix Table A.5; top panel) and the SST fields for each of the 20 CMIP5 models (Appendix A.1; other panels). Red dashed lines indicate the latitudes 25°N and 55°N, area within which was determined the minimum latitudinal SST gradient at each longitude (black dots).

3. To identify the zone that would capture most of the variability between the subtropical and subpolar gyres in the real-world and in all the CMIP5 models, a density plot of the locations at which the most intense cooling occurs in the North Atlantic (i.e. the black dots in Figure 3.14) was generated (Figure 3.15). On the western side of the North Atlantic, the paths of the transition between the subtropical and subpolar regions in the CMIP5 models present good agreement among themselves (with an approximate agreement of 60% across the models) and with the observation-based path (Figure 3.15). However on the eastern side of the basin, the paths of the transition diverge within the CMIP5 models, which broadly do not capture the observation-based path (Figure 3.15). Indeed, most of the CMIP5 models tend to follow a latitudinal path (e.g. the CMCC-CESM, MIROC-ESM-CHEM models in Figure 3.14), while the observation-based path follows a discontinuous path (probably influenced by the bathymetry), which is captured by fewer models (e.g. the GFDL-ESM2G model in Figure 3.14). The definition of a zone of transition that is purely defined by latitudinal bands may therefore not be the most accurate representation of the transition between the subtropical and subpolar regions (an angled zone might provide a more realistic representation of the transition). Nevertheless, since the study of the $fCO_{2\text{-ocean}}$ in the subregions of the North Atlantic is not the major focus of this thesis (the basin-wide North Atlantic will be the main focus on Chapters 4 and 5), the approximation of a latitudinal zone of transition is sufficient in this context. I would argue that the region definition method developed here still provides an improvement from the region boundaries defined in other studies (e.g. Mikaloff Fletcher et al., 2007; Schuster et al., 2013), which separate the subtropical and subpolar regions without accounting for a zone of transition and might therefore provide regional results that are contaminated with the biogeochemical signatures from the neighbouring region.

   Here, the zone capturing most of the transition between the subtropical and subpolar regions in the CMIP5 models and in the real-world is defined from 37°N to 47°N (red lines in Figure 3.15). The subtropical region is therefore defined from 10°N to 37°N and the subpolar region from 47°N to 70°N. In the subsequent $fCO_{2\text{-ocean}}$ regional seasonal analysis, the $fCO_{2\text{-ocean}}$ values within that zone were therefore excluded.

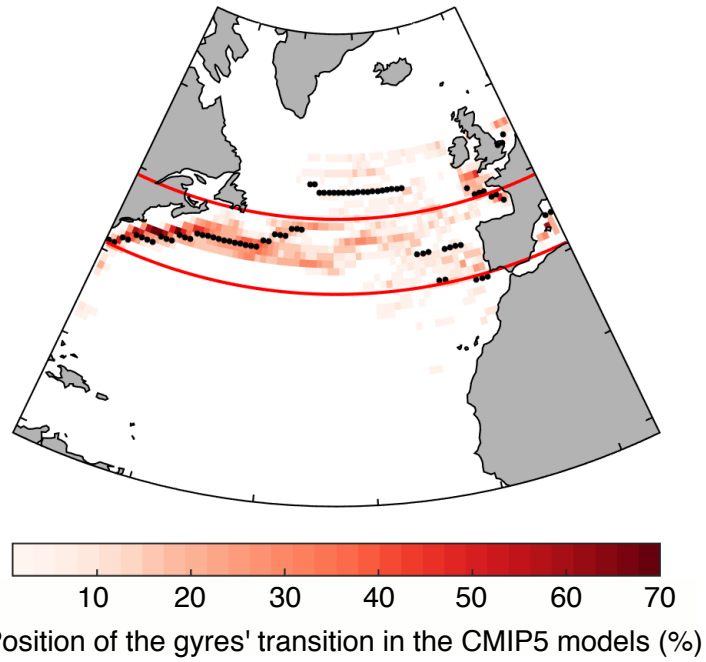

Figure 3.15: Position of the gyres' transition. Density of the position of the transition path between the subtropical and subpolar regions in the CMIP5 models (in %). For example, a grid cell with the value 40% means that 40% of the 20 CMIP5 models (i.e. 8 models) have their maximum northward cooling occurring at the same latitude, for that given longitude. The black dots correspond to the gyres' transition in the real-world (identical to the black dots from top panel in Figure 3.14). The red lines (i.e. 37°N to 47°N) correspond to the selected zone of transition.

In the subtropical region, the CMIP5 models show good agreement with the observed $fCO_{2\text{-ocean}}$ seasonal cycle, but yet tend to overestimate the amplitude of the observed signal (Figure 3.16a - bottom panel). However in the subpolar region, the CMIP5 models are in poor agreement with the observed signal; except for the GFDL-ESM2M and NorESM1-ME models, which capture relatively well both the phase and amplitude of the observed signal (Figure 3.16b - bottom panel). As such, the high variability in the $fCO_{2\text{-residuals}}$ in the subpolar region relative to the subtropical region for most of the CMIP5 models, which was discussed above (Figure 3.11), seems to arise from poor seasonal representation in most of the CMIP5 models, rather than for example, poorly captured annual $fCO_{2\text{-ocean}}$ values. To investigate whether the subtropical and subpolar observed seasonal signals potentially characterise representative biogeochemical features or result from a poor sampling artefacts, the number of $fCO_{2\text{-ocean}}$ values per month and per region is analysed (Figure 3.16 - top panels). Overall, the total number of $fCO_{2\text{-ocean}}$ values is relatively similar between the subtropical and subpolar regions (22,010 and 21,790 respectively). The seasonal distribution

of the number of $fCO_{2\text{-ocean}}$ values is also relatively consistent throughout the climatological year, with over a thousand value recorded for all the months and for both regions (Figure 3.16 - top panels). As such, since the sampling distribution between the regions is relatively similar, the deduced seasonal cycles potentially describe a comparable biogeochemical state, which helps the interpretation of the seasonal signals.

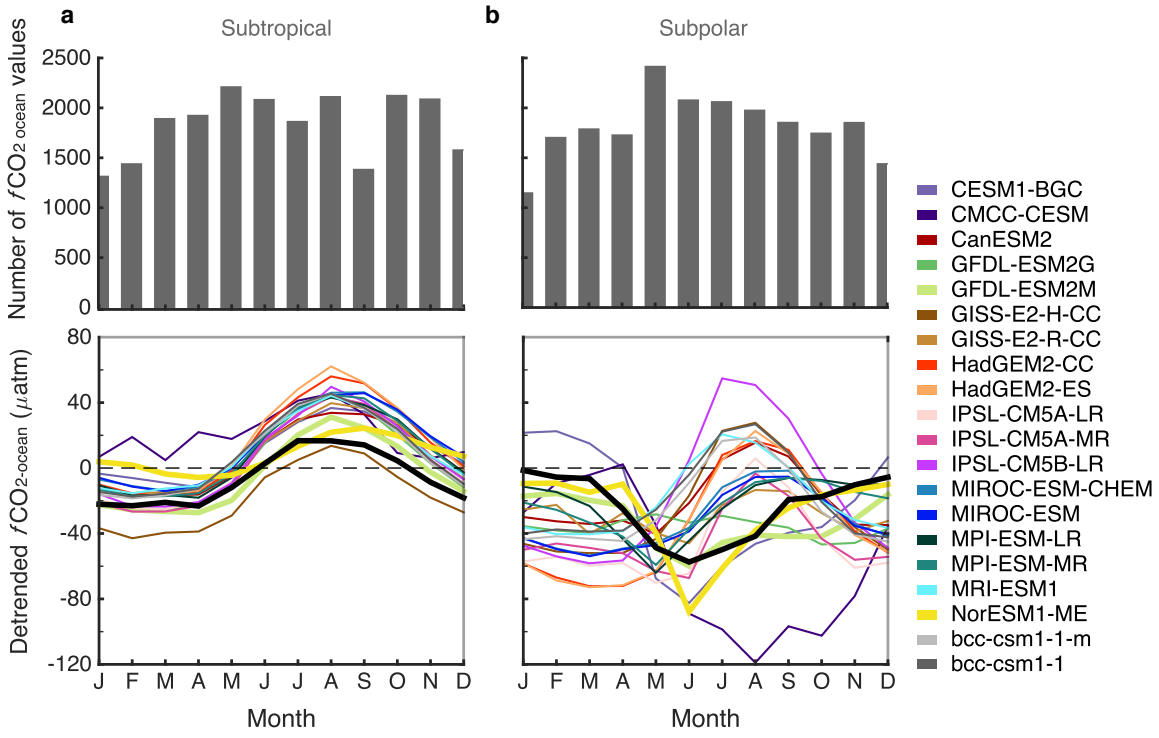

Figure 3.16: The $fCO_{2\text{-ocean}}$ seasonal cycle in the subtropical and subpolar regions of the North Atlantic. **a**, Subtropical (from 10°N to 37°N) and **b**, subpolar (from 47°N to 70°N) analyses on: **top** the number of $fCO_{2\text{-ocean}}$ values within SOCATv4 per month over the period 1970-2015; and **bottom** the detrended seasonal cycle for the SOCATv4 (thick black) and for each CMIP5 models (colours).

The difference between the observed subtropical and subpolar $fCO_{2\text{-ocean}}$ seasonal cycles suggests that the surface biogeochemistry in each region is driven by different regimes. While the subtropical region is mostly temperature-driven, through the impact of the change in SST on the CO_2 solubility in seawater, the subpolar region is driven by a mix of temperature and biological factors (Bennington et al., 2009; Schuster et al., 2013). The poor model skill in the subpolar region is therefore likely to be linked to the models' misrepresentation of the biological blooms, their timing and impact on the surface $fCO_{2\text{-ocean}}$. While the biologically-driven $fCO_{2\text{-ocean}}$ seasonal signal in the models could be due poor representation of the seasonality in the mixed layer depth and hence linked to missing physical mechanisms in the models, fur-

ther investigations (not shown) suggest that the CMIP5 models (1) capture relatively well the seasonal phase of the mixed layer depth in both the subtropical and subpolar regions and (2) seem to poorly represent the relationship between the seasonal SST and the seasonal net primary production in the subpolar region. The fact that the GFDL-ESM2M and NorESM1-ME models capture the subpolar $fCO_{2\text{-ocean}}$ seasonal cycle suggest that their timing and impact of the biological blooms on the $fCO_{2\text{-ocean}}$ seasonality are better represented than the other CMIP5 models. Interestingly, the GFDL-ESM2G model, which uses the same biogeochemical model as the GFDL-ESM2M (i.e. the TOPAZ2 model; Appendix Table A.1), does not present a seasonal model skill as good as the ESM2M version (i.e. the GFDL-ESM2G MAE on the North Atlantic and subpolar region seasonal cycles are 12.2 µatm and 24.5 µatm, respectively, while for the GFDL-ESM2M model are 2.8 µatm and 10.6 µatm, respectively). As such, the different model skills between the two GFDL model versions highlights the sensitivity of the ocean carbon to the different ocean model dynamics (e.g. the ESM2M version uses a z-coordinate ocean model, while the ESM2G provides an isopycnal-coordinate ocean model) (Dunne et al., 2012).

To investigate the importance of the model biases in the subpolar seasonal $fCO_{2\text{-ocean}}$ in capturing the overall North Atlantic seasonal signal, the Mean Absolute Error (MAE) between the models' and the observational seasonal signals are explored (Figure 3.17). A positive correlation between the MAE in the $fCO_{2\text{-ocean}}$ seasonal signal in the subpolar and North Atlantic areas is found (R^2=0.77). As such, models that have a $fCO_{2\text{-ocean}}$ seasonal signal in the subpolar region close to the one in the real-world also tend to have relatively realistic seasonal signal in the North Atlantic (given the observational coverage) (e.g. the NorESM1-ME model). While the causality of the positive relationship between the biases in the subpolar and North Atlantic $fCO_{2\text{-ocean}}$ seasonal signals cannot be guaranteed, the above seasonal analysis suggests for the CMIP5 models to capture the North Atlantic $fCO_{2\text{-ocean}}$ seasonality relatively well, they should be able to describe key interactions between the surface biogeochemistry and the physical and biological processes associated with the spring blooms of the subpolar region.

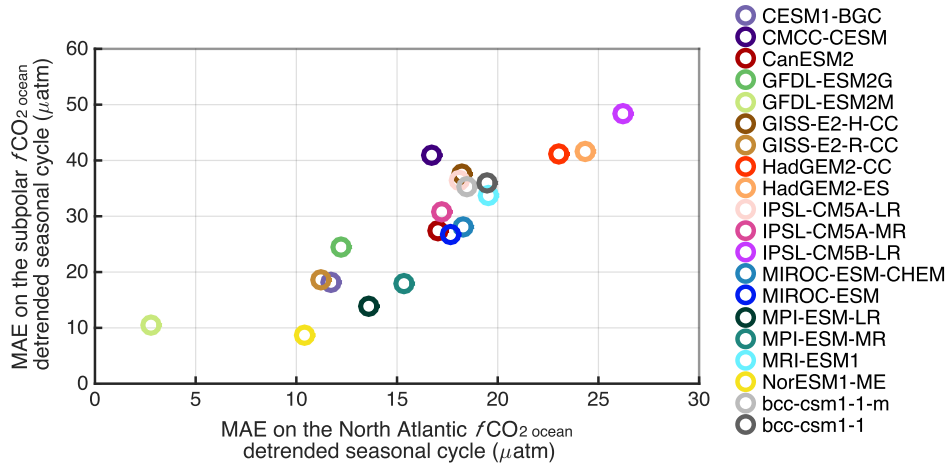

Figure 3.17: The role of subpolar $fCO_{2\text{-ocean}}$ seasonal cycle for capturing the North Atlantic seasonality. Mean Absolute Error (MAE) between the detrended $fCO_{2\text{-ocean}}$ seasonal cycle from each CMIP5 model and the detrended $fCO_{2\text{-ocean}}$ seasonal cycle from SOCATv4, using the seasonality in the subpolar region (y-axis; Figure 3.16b-bottom) and in the North Atlantic (x-axis; Figure 3.13).

3.3 Summary

In summary, the North Atlantic is a basin where the surface CO_2 observational record has been substantially improved and maintained over the past decades. However, remaining gaps in the observational spatial and temporal coverage lead to substantial challenges when robustly determining and interpreting the mean $fCO_{2\text{-ocean}}$ state, and thus the magnitude of the air-sea CO_2 flux. To explore the impact of the current observational sampling scheme on the time-varying surface CO_2 system, models are used as "investigation tools". By providing a perfectly defined world, models allow us to test statistical methods which would be impossible to apply in the real-world with the current observational coverage, such as the quantification of the sampling coverage impact on basin-wide (model-truth) annual means and trends. The model-based study showed that (1) despite the relatively well-developed coverage since ~2005, annual means calculated from data that follow the observational CO_2 coverage underestimate the annual means that would be returned if the North Atlantic was sampled at every grid cell and month, and (2) the large trends' standard error, which is due to the year-to-year sampling variability, prevents the robust determination of whether the North Atlantic sink recently increased or decreased. However, without any further justification, the model-based sampling errors/biases calculated here cannot be directly applied to the real-world. Indeed, such assumption would rely on the CMIP5 models capturing the

observational $fCO_{2\text{-ocean}}$ spatial and temporal variability, which as shown through the model evaluation analyses, is not guaranteed for most models. Overall, the CMIP5 models tend to overestimate the $fCO_{2\text{-ocean}}$ observed variance, particularly in the subpolar region where key interactions between the biology and the surface CO_2 biogeochemistry seem to commonly be missing. This work represents some of the most thorough evaluation of the CMIP5 ocean biogeochemical model results undertaken to date, and brings in to question how well the CMIP5 models capture the observed air-sea CO_2 flux. This chapter therefore highlights the necessity to provide North Atlantic basin-wide and temporally-varying $fCO_{2\text{-ocean}}$ estimates with robust uncertainty so that (1) the change in the surface $fCO_{2\text{-ocean}}$ and hence of the air-sea CO_2 flux in the CMIP5 models can be robustly evaluated, and (2) the real-world response to the increasing atmospheric CO_2 concentration can be determined and better understood. This will be the focus on the next two chapters.

Chapter 4

Quantifying recent change in the North Atlantic CO_2 sink with robust uncertainties

By making use of the strengths of both observational and model data, this chapter aims to robustly quantify basin-wide uncertainties on the North Atlantic annually-varying surface fugacity of CO_2 ($fCO_{2\text{-ocean}}$), and therefore improve our understanding of the recent change in the CO_2 oceanic sink in that basin.

While the large-scale processes controlling atmospheric CO_2 uptake by the North Atlantic are well understood, estimates of the time-varying sink over recent decades range from a decline (Schuster and Watson, 2007), through variability (Gruber et al., 2009; Bates et al., 2014), to an increase (Ullman et al., 2009); Schuster et al., 2013; Landschützer et al., 2016) (Chapter 1). These differences are likely to arise from; the choice of study interval (McKinley et al., 2011; Fay and McKinley, 2013), internal climate variability (McKinley et al., 2011; McKinley et al., 2016), gaps in the observational coverage (Chapter 3), and poor quantification of uncertainty. While various techniques (e.g. Schuster et al., 2013; Landschützer et al., 2013) have been developed to estimate the $fCO_{2\text{-ocean}}$ away from measurement locations and have been widely used to determine the basin-wide air-sea CO_2 fluxes, the time-varying uncertainties associated with such basin-scale interpolations are not well quantified. Here, the North Atlantic averaged $fCO_{2\text{-ocean}}$ from 1992 to 2014 is calculated using a novel observation-based estimate based on a Multiple Linear Regression (MLR) interpolation technique (Watson et al.,

2009), and the uncertainty introduced by which is robustly quantified. For the first time, this basin-wide uncertainty allows us to confidently determine whether the air-sea CO_2 flux trend is significant, and will allow us, in the following chapter, to robustly compare the observation-based results with those simulated by the current generation of ESMs.

This chapter will first describe some of the interpolation techniques that have been developed in the past, highlighting the limitations of the corresponding interpolation uncertainty estimates (Section 4.1). Then, the interpolation method used in this thesis and its corresponding uncertainty assessment are presented (Section 4.2). Once the optimal method for robustly quantifying the uncertainties of the annually time-varying $fCO_{2\text{-ocean}}$ and of the trend over the interval of study is identified (Section 4.3), the recent change in the North Atlantic surface $fCO_{2\text{-ocean}}$ is quantified (Section 4.4). To give confidence in the interpolated $fCO_{2\text{-ocean}}$ estimates and associated uncertainties, a validation assessment is undertaken (Section 4.5), and the limitations behind the spatially-varying uncertainty estimates are discussed (Section 4.6). Finally, the recent change of the North Atlantic CO_2 sink stemming from the method assessment is presented (Section 4.7).

4.1 Interpolation methods: main characteristics and limitations

To overcome the irregular temporal and spatial distribution of *in-situ* sea surface CO_2 measurements (Chapter 3), various statistical techniques based on our current knowledge of the observed CO_2 system have been developed. In this section, three methodologically independent and complementary techniques are presented: the MLR approach (Section 4.1.1), the Neural Network approach (Section 4.1.2) and a statistical gap-filling approach (Section 4.1.3). To highlight the difficulties when determining basin-wide interpolation uncertainty, the uncertainty assessment used in each study presented here, is also discussed.

4.1.1 Multiple Linear Regressions

The MLR approach relies on the mechanisms that link the predicted variable, which is described by the spatially and temporally discontinuous surface $fCO_{2\text{-ocean}}$, to a set of

explanatory variables, which could be any known basin-wide mechanistically-driven variables that are involved in the $fCO_{2\text{-ocean}}$ response to anthropogenic changes, as well as physical and biogeochemical oceanic properties (e.g. SST, wind speed, chlorophyll-a, atmospheric CO_2 concentrations). By building a linear relationship between the predicted and the explanatory variables from the locations and times at which the $fCO_{2\text{-ocean}}$ observations are available, the deduced regression coefficients are applied to the basin-wide and continuous explanatory variables, which enables to fill the gaps in the surface $fCO_{2\text{-ocean}}$ (Figure 4.1).

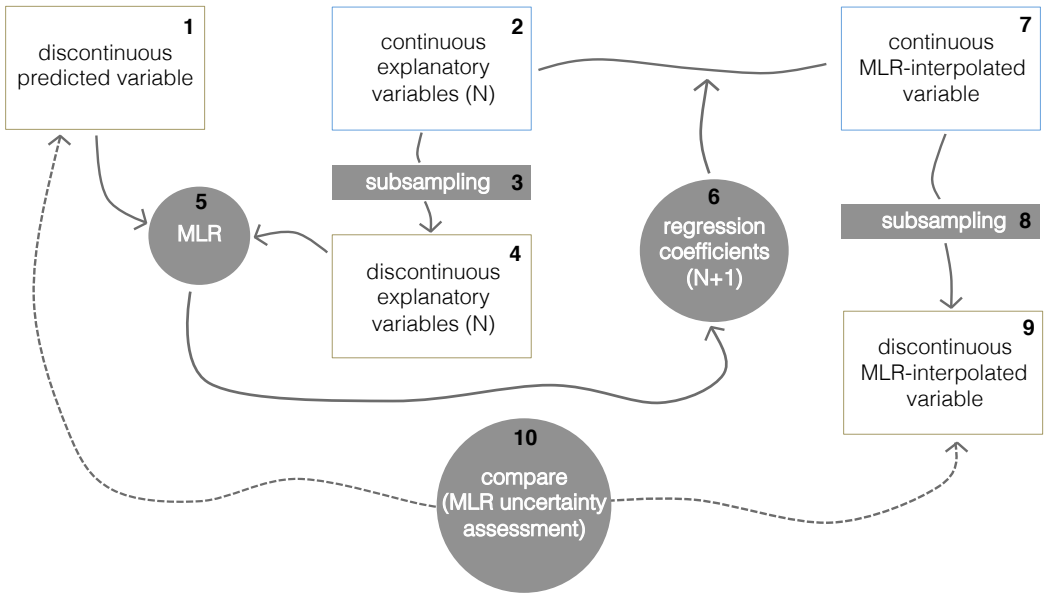

Figure 4.1: The general MLR approach. Schematic representation of the main steps involved in a MLR approach used to interpolate a predicted variable (**1**) to a basin-wide region and at the frequency of interest (**7**), where brown and blue boxes refer to the discontinuous and continuous data, respectively, and where filled items correspond to data-processing steps. The "subsampling" step refers to the extraction of a continuous field following the discontinuous observational coverage in the predicted variable. "N" indicates that N explanatory variables can be used in a MLR approach, and consequently N+1 regression coefficients are deduced, where the (+1) corresponds to the intercept. The MLR uncertainty assessment is achieved by comparing, at places and times at which observations were made, the MLR-predicted field (**9**) with its original values (**1**).

To account for specific relationships between the predicted and explanatory variables across different biogeochemical regimes (e.g. in the North Atlantic subtropical gyre, the surface $fCO_{2\text{-ocean}}$ is mostly temperature-driven, while in the subpolar region, it is principally temperature and biologically-driven; Schuster and Watson, 2007; Chapters 1, 3), MLR-based studies generally build a linear fit within each geographical regime, whose

separated $fCO_{2\text{-ocean}}$ results are finally gathered to reconstruct the basin of interest (e.g. Iida et al., 2015). Similarly, to potentially improve the statistical model's ability to capture specific temporal features, the period of interest can be divided into shorter intervals, over which separated linear fits are performed. For instance, to study the air-sea CO_2 flux in the North Atlantic in 2005, Watson et al. (2009) performed a MLR between the observed $fCO_{2\text{-ocean}}$ and three explanatory variables (i.e. SST, Mixed Layer Depth (MLD) and geographical location) for each discrete latitude bands of 30°, 20° and 10° width and for each 2-month interval. While the division of the space and time dimensions enables the specific relationships between the predicted and explanatory variables to be captured, two major drawbacks can arise from subdivision into regions and periods which individually contain little data: (1) the amount of information used by the MLR might become too limited to produce robust linear fit (i.e. the system is underdetermined), and (2) the linear model might capture specific features at the site of individual measurements, but may not be applicable to the wider areas and periods. It is therefore crucial to find an appropriate balance between feeding the MLR with data that would capture specific features of variability and giving enough data to produce a MLR with good predictive skill. Also, one should be aware of that overfitting can arise when providing too many explanatory variables to the MLR or when letting the statistical linear model finding the best fit across the possible combinations within the given explanatory variables (Hastie et al., 2016, Chapter 7). If overfitting occurs, the resulting regression coefficients will manage to capture the noise in the predicted variable rather than describing the overall population. As such, while the skill of the linear fit would present statistically good values (e.g. high correlation between the predicted field and the original values), the relationship would most likely not fit any new data and therefore acts as a poor predictive model.

To assess the performance of a linear model, the MLR-predicted field is commonly compared to its original observational values (Figure 4.1), and as such the uncertainty assessment is limited to the locations and times at which the measurements were conducted. Often, the interpolation uncertainty is quantified by the RMSE between the MLR-predicted and the observational data (e.g. Schuster et al., 2013; Iida et al., 2015). Substantial assumptions are therefore made when applying the "localised" uncertainty to the basin-wide and continuous

predicted field; one of the assumption being that the interpolation method does not add any bias on the time-varying surface $fCO_{2\text{-ocean}}$. To evaluate (to a first degree) the predictive skill of a linear model, bootstrapping analyses can be undertaken. In a bootstrapping analysis, a subset of the initial training dataset is removed and the remaining data are used to perform a new MLR. The predicted $fCO_{2\text{-ocean}}$ from the new MLR can then be compared to the removed subset of $fCO_{2\text{-ocean}}$ observations. By generating a bootstrapping analysis successively to different removed subsets, the predictive skill of the different new MLRs can be evaluated, but is still ultimately constrained to the discontinuous observational coverage and to sets of data that may be dependent (e.g. data collected from the same voyage), and does not guarantee that the deduced interpolation uncertainty reflects the predictive skill of the MLR that uses all the available information (i.e. no removed data).

To assess the basin-wide uncertainty on the MLR-predicted $fCO_{2\text{-ocean}}$ values, Watson et al. (2009) used two methods. The first method was based on empirical semivariograms, which are geostatistical tools that specifically evaluated the spatial continuity and variability of the $fCO_{2\text{-ocean}}$ residuals (i.e. the difference between the MLR-predicted and the observational $fCO_{2\text{-ocean}}$ values) across the North Atlantic basin (Webster and Oliver, 2007). The second method was based on the $fCO_{2\text{-ocean}}$ output of a high resolution (i.e. 1/12° degree) biogeochemical model, which provides the main advantage of being known at each grid cell and at each time step. Specifically, the model $fCO_{2\text{-ocean}}$ field was subsampled at the same locations and days of the year 2005 (i.e. the year of study) as the observational coverage and was then treated as real observations. The subsampled $fCO_{2\text{-ocean}}$ model field was defined as the predicted variable and the observed SST, MLD and geographical position as the explanatory variables. The MLR-predicted $fCO_{2\text{-ocean}}$ was then compared to the model-truth $fCO_{2\text{-ocean}}$ values and whose comparison was therefore achieved at the North Atlantic basin-wide scale and over each time step. While the second approach is particularly useful to evaluate the basin-wide uncertainties and test for potential biases introduced by the interpolation procedures, fundamental issues arise from this specific study. Indeed, by using observational-based data for the explanatory variables instead of the model's SST and MLD, which would have returned a model-coherent study, Watson et al. (2009) relies on the fact that the variability in the modelled surface $fCO_{2\text{-ocean}}$ perfectly matches the variability of the observed physical field. Such assumptions are particularly difficult to verify, because: (1) the biogeochemical

model was forced using climatological surface fields (i.e. wind-stress and heat fluxes), which likely do not represent the year 2005 (whether the climatologies were normalised to 2005 is not indicated), (2), the model $fCO_{2\text{-atmosphere}}$ was held at pre-industrial values, which differs from what the physical explanatory variables have been experiencing in the real-world (Chapter 1), and (3) even if the biogeochemical model were to represent the year 2005, the uncertainty assessment relies on one specific model analysis, which might be limited to fully capture the variability of the interpolation procedure. The MLR approach and the uncertainty assessment developed by Watson et al. (2009) will be extended and improved later in the chapter (Sections 4.2; 4.5).

4.1.2 Neural networks

Instead of conventionally using fixed boundaries to define the basins over which an interpolation method is applied, other methods, such as the neural network approach, enable the definition of biogeochemical provinces whose boundaries change from one time step to the next (Longhurst, 2007; Landschützer et al., 2013). The neural network approach has been largely developed within the ocean biogeochemistry community with the work of Landschützer et al. (2013). In their study, Landschützer et al. (2013) fed a neural network algorithm with physical and chemical proxies (i.e. SST, Sea Surface Salinity (SSS) and MLD) that describe the surface $pCO_{2\text{-ocean}}$ monthly climatology (Takahashi et al., 2009). By detecting common features within the input datasets, the neural network approach can identify clusters with similar biogeochemical behaviour; each of them defining a specific biogeochemical province. In regions of strong seasonal $pCO_{2\text{-ocean}}$ variability (e.g. the subpolar North Atlantic), the grid cells are likely to belong to various biogeochemical provinces (Landschützer et al., 2014). Once the biogeochemical provinces defined for each step of the period of study, Landschützer et al. (2013) then reconstructed the surface $pCO_{2\text{-ocean}}$ within each province using an interpolation approach similar to a MLR technique (i.e. the feed-forward network).

While the neural network approach enables the capturing of geographical variability in marine biogeochemistry, few aspects of such approach require further consideration: (1) waters from different ocean basins can belong to the same biogeochemical cluster and be treated similarly by the interpolation method (e.g. some surface waters near the Bering Strait, the Labrador Sea and in some areas of the Southern Ocean all belong to the same province in

Landschützer et al. (2014) and are therefore treated equally), (2) the neural network relies on the monthly climatological $pCO_{2\text{-ocean}}$ (Takahashi et al., 2009) being bias-free with no source of uncertainty, which is not necessarily the case due to the substantial gaps in the observational coverage, particularly in the Southern Hemisphere, and (3) the neural network learns from a monthly climatological $pCO_{2\text{-ocean}}$ field in relation with the time-varying SST, SSS and MLD fields, which might perturb the detection of specific clusters that are linked to internal variability or climate change, that the climatological field likely does not capture.
To validate the neural network-based $pCO_{2\text{-ocean}}$ estimates, Landschützer et al. (2013) directly compared the predicted results with the observations and quantified the interpolation uncertainty by calculating the RMSE between the two $pCO_{2\text{-ocean}}$ fields. To assess the predictive skill of the interpolation method, Landschützer et al. (2013) compared the neural network-based results against independent observations (i.e. observations that were not used to train the feed-forward network algorithm). For instance, they used the newly stored $pCO_{2\text{-ocean}}$ observations in SOCAT version 2 (the interpolation method was trained on the previous version of SOCAT, the version 1.5) and compared them with the predicted $pCO_{2\text{-ocean}}$ values. By using these external observations, Landschützer et al. (2013) were able to identify a larger RMSE and bias than when using the training observations only (i.e. respectively $\sim$23 µatm against 10 µatm for the RMSE, and 4.85 µatm and -0.10 µatm for the bias). While using external observations is useful to obtain an overall assessment of the predictive nature of the interpolation method outside of the training dataset, it still remains impossible with spatially and/or temporally sparse observations to determine the basin-wide and time-varying uncertainty due to the interpolation method.

4.1.3 Statistical gap-filling approach

Unlike the MLR and neural network approaches that rely on the mechanisms linking the $fCO_{2\text{-ocean}}$ (or $pCO_{2\text{-ocean}}$) to physical, biological and/or chemical variables, the statistical gap-filling approach developed by Jones et al. (2015) only relies on the spatiotemporal variability of the observed $fCO_{2\text{-ocean}}$ field. While such an approach prevents the intrusion of potentially poorly represented relationships between surface $fCO_{2\text{-ocean}}$ and proxy variables, and of additional sources of uncertainties contained in the proxy variables (e.g. uncertainty on the measurements, uncertainty on the data-assimilation technique for some MLD products), the

Jones et al. (2015) statistical gap-filling approach relies on the CO_2 observational coverage capturing most of the spatial and temporal variability and sampling localised and punctual events of extreme variability. As such, the Jones et al. (2015) statistical gap-filling approach likely omits or misrepresents substantial $fCO_{2\text{-ocean}}$ features in areas of limited measurements, especially in areas of intense variability (if sampled in the first place), such as the Equatorial Pacific. The method developed by Jones et al. (2015) is described by four main statistical procedures that are based on temporal and spatial interpolation techniques, which were applied to each grid cell of the global ocean (south of 70°N):

1. A temporal curve fit was applied to the $fCO_{2\text{-ocean}}$ observations available in the grid cell.

2. If the temporal fit was found to be unrealistic because it did not follow certain statistical criteria (e.g. in order to avoid overfitting, the fitted curve should contain one dominant plankton bloom per year) and/or because too few (or no) measurements were available within the studied grid cell, a spatial interpolation with the neighbouring grid cells was achieved via an autocorrelation method until the curve fit was realistic in terms of fitting the observations and the relatively well-known seasonal characteristics of the surface $fCO_{2\text{-ocean}}$.

3. At times when original $fCO_{2\text{-ocean}}$ observations were available, the $fCO_{2\text{-ocean}}$ curve fitted values were replaced by the observations.

4. Discontinuities arising from the combination of temporal curve fitting, spatial interpolation and observational adjustments were finally attenuated with a "smooth spline" function.

Unlike most interpolation techniques that neglect the uncertainty on the $fCO_{2\text{-ocean}}$ observations, Jones et al. (2015) initiated their uncertainty assessment with a fixed measurement uncertainty of $\pm$ 2.5 µatm (Takahashi and Sutherland, 2013), which then propagated through the system with the uncertainty associated on the curve fit, autocorrelation and interpolation methods. As such, Jones et al. (2015) provided a spatially and temporally complete uncertainty on the gap-filled $fCO_{2\text{-ocean}}$, but whose large amplitude limits the value of this approach when attempting to identify relatively small trends. Indeed, Jones et al. (2015) identified that over the period 1985-2011, the difference between the global ocean trend and the atmospheric

trend was -0.18 $\pm$ 0.76 µatm·yr^{-1}; where a negative difference indicates that the trend in the $fCO_{2\text{-ocean}}$ is smaller than the trend in the $fCO_{2\text{-atmosphere}}$, and that the air-sea CO_2 flux would increase with time in the absence of significant trends in the SST, SST and/or surface winds (Equation 1.4). Since the uncertainty (i.e. 0.76 µatm·yr^{-1}) is about four time larger than the amplitude of the mean trend difference (-0.18 µatm·yr^{-1}), it therefore remains unclear using this gap-filling interpolated product whether the global oceanic CO_2 sink increased or decreased over the past decades. Nevertheless, such a result might also be impacted by the method used to calculate the trend. The trends in the $fCO_{2\text{-ocean}}$ and $fCO_{2\text{-atmosphere}}$ were calculated as the difference between the 1985-1989 mean and the 2007-2011 mean, divided by the length of the product record (i.e 27 years). As such, Jones et al. (2015) might (for example) have been sampling two periods of oceanic internal variability and therefore might not capture the overall response of the surface $fCO_{2\text{-ocean}}$ system due to the increase of CO_2 concentrations in the atmosphere.

4.2 An interpolation method with a novel uncertainty assessment

While various interpolation methods exist in the literature, basin-wide and time-varying interpolation uncertainties either do not exist, or are too broad to allow the detection of significant trends in surface $fCO_{2\text{-ocean}}$ and therefore in the CO_2 oceanic sink (Section 4.1). This limits our understanding of the marine carbon response due to increasing atmospheric CO_2 and climate change, but also limits the evaluation and investigation of CO_2 uptake simulated by climate models. In this section, the overall structure of the interpolation technique with its associated uncertainty assessment that is used for the rest of the chapter, and whose results will be exploited in the following chapter, is first presented (Section 4.2.1). Then the data processing necessary for the generation of interpolated surface $fCO_{2\text{-ocean}}$ with basin-wide uncertainties is introduced (Section 4.2.2), followed by a more detailed explanation of the various steps involved in our method (Section 4.2.3).

4.2.1 General procedures

The MLR method used throughout the rest of this thesis is an extension of Watson et al. (2009) (Section 4.1). While more recent MLR methods have been published since the work of Watson et al. (2009) (e.g. Schuster et al., 2013; Iida et al., 2015), the specific focus on Watson et al. (2009) is justified by the fact that (1) Watson et al. (2009) assessed for the first time the basin-wide North Atlantic uncertainty associated with the interpolation method (the determination of robust basin-wide uncertainties on the North Atlantic CO_2 sink was initially the main motivation of this Ph.D.), and (2) the lead author (A.J. Watson) is a supervisor on this Ph.D., allowing me to thoroughly interrogate him about decisions made in his work, and the leader of the project to which this Ph.D. is affiliated (i.e. the RAGNARoCC project). In their study, Watson et al. (2009) built, for each 2-month interval of the year 2005 and for subdivisions of the North Atlantic of 30°, 20° and 10° latitude, a linear relationship between the surface $fCO_{2\text{-ocean}}$ and three explanatory variables: the SST, MLD and longitude. By being directly linked to changes on CO_2 concentration in seawater, the SST and MLD are characterised as mechanistically-driven explanatory variables. Indeed, the solubility of CO_2 increases as the temperature of the surface waters decreases, and the deepening of the MLD through density-influenced and/or wind-mixing events can (1) enhance the dilution within the mixed layer of additional CO_2 taken up by the surface ocean and hence stimulate further uptake, (2) bring nutrient-enriched waters to the euphotic layer, potentially enhancing photosynthesis and the CO_2 uptake, and (3) also bring DIC-enriched waters to the surface, leading to local CO_2 outgassing (c.f. Chapter 1). Note that Watson et al. (2009) initially included chlorophyll-a as an additional explanatory variable, which would provide a better proxy for the impact of biological activity on surface CO_2 concentration than MLD, but was finally discarded due to (1) non-significant improvements on the surface $fCO_{2\text{-ocean}}$ predictions, and (2) the spatial and temporal limitations of the chlorophyll-a product, which constrained the prediction of surface $fCO_{2\text{-ocean}}$ to areas and periods for which data from the satellite-derived product was available. The longitude explanatory variable included in the Watson et al. (2009) MLR method acts as a "statistically-driven" variable, aiming to account for east-west biases that may exist in the observational coverage, and hence improve the prediction. While this thesis could have been focused on investigating the optimal configuration of explanatory variables to provide the best basin-wide surface $fCO_{2\text{-ocean}}$ estimates, which

is an on-going work carried out by U. Schuster and generally widely-investigated within the ocean biogeochemistry community, the commitment made for this Ph.D. was to gain substantial new understanding from what established and widely-used mapping approaches could tell us about the North Atlantic CO_2 uptake. Nevertheless, the method used in Watson et al. (2009) is here extended by:

1. optimising latitude bands extent to 60° (which corresponds to the width of the North Atlantic[1]), 5°, 2° and 1° in order to account for the approximately latitudinally separated biogeochemical regimes (e.g. subtropical gyre convergence, Gulf Stream transport, subpolar gyre divergence) due to ocean circulation, but also to investigate the impact of the spatial division on the predicted $fCO_{2\text{-ocean}}$,

2. widening the period of study from a single year (i.e. 2005) to the time-varying period 1992-2014 (the linear relationship was tested from 1992 as it corresponds to the starting year of the MLD assimilation product; Table A.5),

3. incorporating the atmospheric CO_2 mixing ratio (xCO_2) as a fourth explanatory variable in order to account for the time-varying impact of anthropogenic CO_2 emissions on the marine carbon system,

4. providing a more comprehensive assessment of the basin-wide time-varying uncertainty. The MLR analysis performed on observational products is hereinafter referred as the "observation-based MLR".

To quantify the basin-wide and time-varying uncertainty associated with the interpolation technique, 19 CMIP5 models are used[2] (Tables A.3, A.1). In this chapter, the CMIP5 models are used as "investigation tools", which by being perfectly known worlds offer an alternative to test how well the interpolation method performs, especially at places where no data is available. Unlike Watson et al. (2009) which assessed the interpolation uncertainty by training a MLR on a biogeochemical-modelled $fCO_{2\text{-ocean}}$ field with observational-based explanatory variables (a first attempt that unrealistically relies on the fact that the model captures the real-world variability; Section 4.1), the present analysis

[1]The North Atlantic is defined from 10°N to 70°N and from -75°E to 5°E.

[2]Based on the $fCO_{2\text{-ocean}}$ model evaluation completed in Chapter 3, the CMCC-CESM model was discarded from the present MLR analysis due to unrealistic variability; leading to a remaining of 19 CMIP5 models out of the initially 20 models available.

generates a MLR that is consistent to the modelled fields: each MLR is built using the $fCO_{2\text{-ocean}}$ field from one of the 19 CMIP5 models and the SST, MLD, atmospheric xCO_2 from the corresponding CMIP5 model (and longitude, which is not a model-specific variable), referring hereinafter to the "CMIP5-based MLR" (Figure 4.2). As such, the MLR-predicted $fCO_{2\text{-ocean}}$ returned by the CMIP5-based MLR analyses are therefore used to determine the ability of the interpolation technique to capture existing features in the model-truth field (Figure 4.2), such as trends.

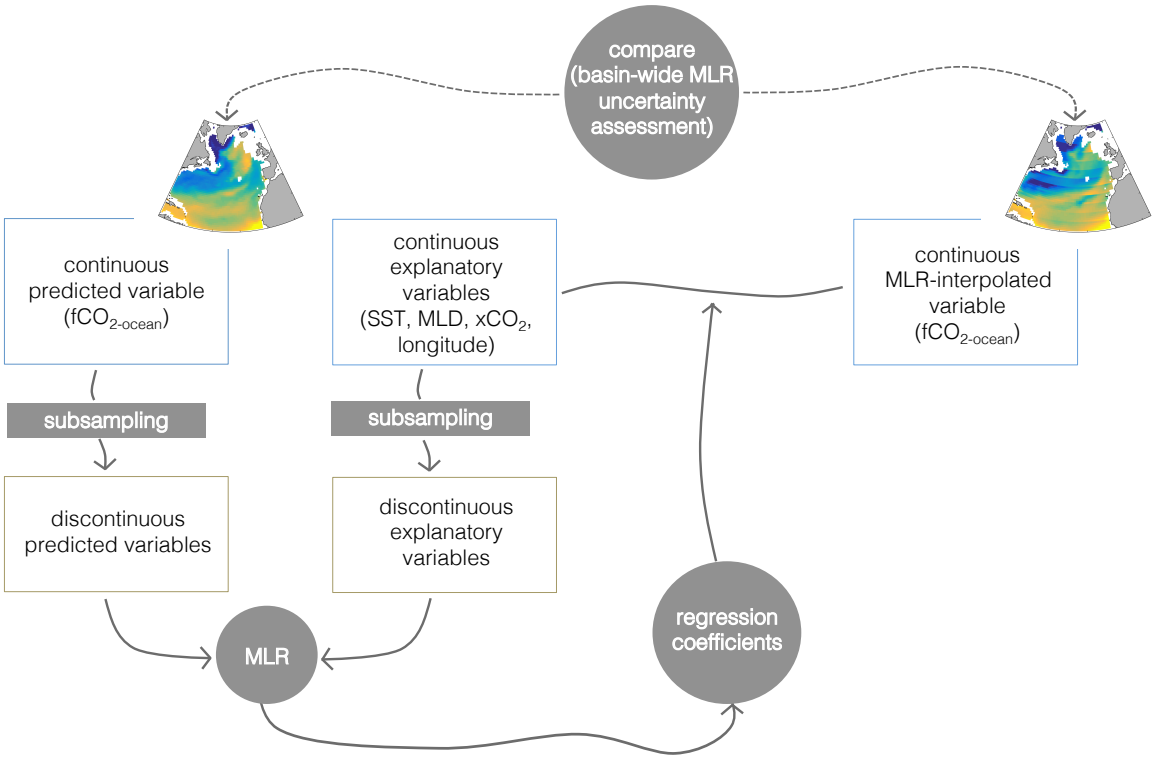

Figure 4.2: The general method to determine basin-wide interpolation uncertainties, based on the CMIP5-based MLRs. Schematic representation of the main steps involved in each CMIP5-based MLR, where brown and blue boxes refer to the discontinuous and continuous data, respectively, and where filled items correspond to data-processing steps. The "subsampling" step refers to the extraction of the continuous modelled field following the discontinuous observational coverage in SOCATv4. Unlike the observation-based MLR (Figure 4.1), the MLR uncertainty assessment in the CMIP5-based MLR is achieved by comparing, at all ocean grid cell and times, the MLR-predicted field (right map) with the original modelled values (left map).

4.2.2 General data processing

For the observation-based MLR, the surface $fCO_{2\text{-ocean}}$ (i.e. the predicted variable) was taken from the SOCATv4 monthly gridded product (Bakker et al., 2014; Chapter 3), and the ex-

planatory variables were taken from the OISST version 2 product (Reynolds et al., 2007) for SST, the Circulation and Climate of the Ocean Phase II (ECCO2) product (Menemenlis et al., 2008) for MLD, and the GLOBALVIEW-CO_2 reference matrix (GLOBALVIEW-CO2, 2013) for xCO_2 (further description on the observational data processing is provided in Table A.5). For the CMIP5-based MLR, the data processing (e.g. data download, conversion of the modelled $pCO_{2\text{-ocean}}$ into $fCO_{2\text{-ocean}}$, calculation of the MLD, regridding into a regular $1^\circ \times 1^\circ$ resolution) is fully detailed in Appendix A.1. Treating the models as plausible alternative worlds, the $fCO_{2\text{-ocean}}$ in each CMIP5 model was subsampled in the same latitude, longitude and date (month and year) as each data point in the SOCATv4 monthly gridded product, creating a parallel "subsampled" modelled dataset (Figure 4.2).

In each observational and CMIP5 modelled product, (1) the freshwater inputs from rivers and coastal effects were minimised by removing regions shallower than 1,000 m water depth (as in Watson et al., 2009), using the Earth Topography (ETOPO)1 Bedrock product (Amante and Eakins, 2015), and (2) the surface fields were extracted for the period 1992-2014 and for the North Atlantic (i.e. from 10°N to 70°N and from -75°E to 5°E.).

4.2.3 The MLR steps

The following steps were identically undertaken for the observation-based MLR and for each of the 19 CMIP5-based MLR analysis, but with their respective variables (e.g. the term SST below refers to the OISST product for the observation-based MLR and to the SST output of a CMIP5 model for the CMIP5-based MLR). Since the CMIP5 model simulations are conditioned with global annual atmospheric xCO_2 values (Chapter 1), and not with seasonally and spatially varying values as in the real-world, the CMIP5-based study is built, for consistency purposes, on model-relevant variables (i.e. choosing annual atmospheric xCO_2 for CMIP5-based MLR and choosing the seasonally and spatially varying xCO_2 for the observation-based MLR).

1. The explanatory variables were normalised, using:

$$\mathrm{Y}| = \frac{\mathrm{Y} - \min(\mathrm{Y})}{\max(\mathrm{Y}) - \min(\mathrm{Y})} \tag{4.1}$$

where Y is an explanatory variable. The normalisation step is particularly useful to

compare the amplitude of the regression coefficients β and therefore identify the variable that dominates the $fCO_{2\text{-ocean}}$ signal (assuming that the explanatory variables are independent).

2. The normalised explanatory variables were subsampled in time and space where $fCO_{2\text{-ocean}}$ data is available in SOCATv4, referred as the "subsampled" data.

3. The MLR analysis was performed using the monthly $fCO_{2\text{-ocean}}$ from 1992 to 2014, in the open waters of the North Atlantic, with the monthly-normalised subsampled SST, MLD, xCO_2 and longitude:

$$fCO_{2-\text{ocean}} = \beta_0 + \beta_1 \cdot \text{SST}|_{\text{subsampled}} + \beta_2 \cdot \text{MLD}|_{\text{subsampled}} + \beta_3 \cdot xCO_2|_{\text{subsampled}} + \beta_4 \cdot \text{Lon}|_{\text{subsampled}} \tag{4.2}$$

where β_0 is the intercept and $\beta_{1,2,3,4}$ are the regression coefficients.

4. The β coefficients were used with the monthly basin-wide normalised explanatory variables, to predict the monthly basin-wide $fCO_{2\text{-ocean}}$ from 1992 to 2014:

$$fCO_{2-\text{ocean,MLR}-\text{predicted}} = \beta_0 + \beta_1 \cdot \text{SST}| + \beta_2 \cdot \text{MLD}| + \beta_3 \cdot xCO_2| + \beta_4 \cdot \text{Lon}| \tag{4.3}$$

In this case (i.e. one MLR performed on the North Atlantic), there is one set of β coefficients. The MLR analysis was repeated by dividing the North Atlantic into 30°, 20°, 10°, 5°, 2° and 1° latitude bands, which returned 2, 3, 6, 12, 30 and 60 sets of regression coefficients, respectively. For instance, when considering the MLR over latitude bands of 20° width, three different MLRs were computed: one using all the SOCATv4 gridded values available (with the corresponding subsampled explanatory variables) within the region 10°N to 30°N, a second one based on region 30°N to 50°N and a final one over the region 50°N to 70°N (all three using all of the North Atlantic longitudes and the data stored in the period 1992-2014). The three sets of regression coefficients are then applied to the corresponding latitude band-wide explanatory variables to predict the latitude band-wide $fCO_{2\text{-ocean}}$ over the same interval, which when gathered, reconstruct the North Atlantic basin-wide $fCO_{2\text{-ocean}}$. For the seven different

spatial MLRs (latitude bands of 60°, 30°, 20°, 10°, 5°, 2° and 1° width), seven different monthly basin-wide $fCO_{2\text{-ocean}}$ estimates for the North Atlantic from 1992 to 2014 were obtained.

In summary, a total of 140 $fCO_{2\text{-ocean}}$ predicted fields over the North Atlantic and covering the period 1992-2014 at a monthly frequency were generated from the various spatially-divided observation-based and CMIP5-based MLR analyses. To identify the optimal spatial MLR analysis and therefore use the corresponding results to quantify the recent change in the North Atlantic surface $fCO_{2\text{-ocean}}$, further investigations using the CMIP5-based MLR results is needed.

4.3 Investigating the optimal MLR method

The reconstructed basin-wide $fCO_{2\text{-ocean}}$ from the CMIP5-based MLR is compared, at every point in space and time, to the model-truth $fCO_{2\text{-ocean}}$ calculated interactively within each of the CMIP5 models. The study of the basin-wide and temporally-varying $fCO_{2\text{-residuals}}$ (i.e. $fCO_{2\text{-residuals}} = fCO_{2\text{-ocean, MLR-predicted}} - fCO_{2\text{-ocean}}$) therefore allows one to answer the questions: Does the MLR method introduce a significant bias on the predicted $fCO_{2\text{-ocean}}$ North Atlantic annual means and on the 1992-2014 linear trend? Which of the spatially-divided North Atlantic MLR methods (i.e. the MLR trained over latitude bands of 60°, 30°, 20°, 10°, 5°, 2° or 1° width) provides the best predictive skill? These questions will be tackled here, with a study on the annually-varying $fCO_{2\text{-residuals}}$ in Section 4.3.1 and on the linear trends in Section 4.3.2, identifying in both cases the optimal MLR method(s). The annually-varying and trend uncertainties are particularly investigated here as they (1) are directly linked to policy questions surrounding carbon budgets and potential saturation of the sinks, (2) have been so far poorly quantified within the community, and (3) are the largest source of uncertainties when determining the recent change in the CO_2 ocean sink (Iida et al., 2015), but also as they will serve the main motivation for the study carried in Chapter 5.

4.3.1 Annually-varying uncertainty

To study whether the MLR method introduces a significant bias on the predicted annual $fCO_{2\text{-ocean}}$ in the North Atlantic and over the period 1992-2014, the annual time series of

the $fCO_{2\text{-residuals}}$ produced by each CMIP5-based MLR analysis is first calculated. Then, the annually-varying model mean (i.e. thick black line in Figure 4.3) and standard deviation σ (i.e. shadings in Figure 4.3) errors are calculated respectively by:

$$\mathrm{ME_y} = \sum_{m=1}^{19} \frac{\overline{fCO_{2-\mathrm{residual,y,m}}}}{19} \tag{4.4}$$

$$\sigma(\mathrm{ME_y}) = \sqrt{\sum_{m=1}^{19} \frac{\left(\overline{fCO_{2-\mathrm{residual,y,m}}} - \mathrm{ME_y}\right)^2}{18}} \tag{4.5}$$

where the horizontal line represents the North Atlantic area-weighted average, *y* corresponds to the year (from 1992 to 2014) and *m* to the CMIP5-based MLR analysis.

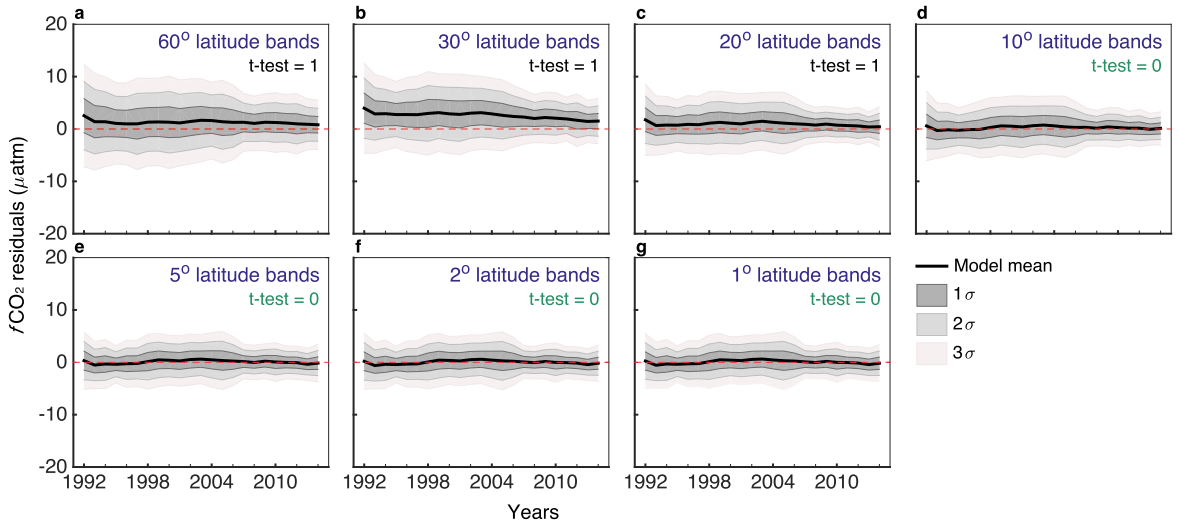

Figure 4.3: Time-varying uncertainty from the CMIP5-based MLR analyses. Each panel corresponds to the results from one of the seven spatially-divided MLR method. The multi-model mean of annual average $fCO_{2\text{-residuals}}$ (black line; Equation 4.4). The dark, medium and light grey shadings correspond, respectively, to the 1, 2 and 3σ of the residuals across the 19 annual averages of $fCO_{2\text{-residuals}}$ (Equation 4.5). The dashed red line indicated the zero level. The t-test result is indicated in each panel, where 0 means that there is no statistically significant mean bias in the residuals, at the 5% significance level.

The annual bias study (Figure 4.3) shows that (1) the smaller the width of the latitude land over which the MLR is trained, the smaller the width of the uncertainty (i.e. the grey shadings in Figure 4.3 are much wider for the 60° study and for the 5° or 1° studies), (2) the uncertainty is smaller toward the end of the period of study (from 2008) than at the beginning, suggesting that MLR is impacted by the increase in number of observations across the North

Atlantic over time, (3) the MLR analyses built on latitude bands of 5°, 2° and 1° width seem to present similar results, suggesting that those three statistical methods provide a similar predictive skill on annual means. A t-test is used to identify whether the uncertainty over the period 1992-2014 is significantly different from a distribution of mean zero at the 5% significance level. The MLR analyses computed successively over latitude bands of 10°, 5°, 2° and 1° width reproduce the model-averaged $fCO_{2\text{-ocean}}$ time series without the addition of a significant bias (i.e. the black thick line is statistically indistinguishable from the zero red line), at the 5% significance level (Figure 4.3d-g). However, for the other analyses (i.e. the ones built over latitude bands of 60°, 30° and 20° width), the linear model overestimates (i.e. the model-mean $fCO_{2\text{-residuals}}$ is positive) the model-true annual $fCO_{2\text{-ocean}}$ values at the 5% significance level (Figure 4.3a-c). As such, the uncertainty assessment from the 10°, 5°, 2° and 1° MLR analyses are potentially all suitable to delimit the annually-varying surface $fCO_{2\text{-ocean}}$ determined from the observation-based MLR (further justification is provided in Section 4.5).

4.3.2 Trend uncertainty

The interpolation uncertainty on the North Atlantic linear trend in surface $fCO_{2\text{-ocean}}$ over the period 1992-2014 is studied by comparing, for each MLR analysis, the $fCO_{2\text{-ocean}}$ trend predicted by the MLR (i.e. $\Gamma_{\text{MLR-predicted}}$) with the model-truth $fCO_{2\text{-ocean}}$ (i.e. $\Gamma_{\text{model-truth}}$) trend given by the model's outputs, across the 19 CMIP5 models (Figure 4.4). For each of the seven different spatial MLR analyses (i.e. 60°, 30°, 20°, 10°, 5°, 2° and 1° latitude bands), the R-squared and the RMSE (Equation 4.6) between $\Gamma_{\text{MLR-predicted}}$ and $\Gamma_{\text{model-truth}}$ were calculated (their values are displayed in each panel of Figure 4.4).

$$\mathrm{RMSE_s} = \sqrt{\sum_{m=1}^{19} \frac{\left(\Gamma_{\mathrm{MLR-predicted,m,s}} - \Gamma_{\mathrm{model-truth,m}}\right)^2}{19}} \tag{4.6}$$

where the indices *s* and *m* correspond to the seven different spatial MLR analyses and to the 19 different CMIP5 models, respectively.

Across the seven spatial MLR analyses, the 5° latitude bands analysis provides the highest R-squared value (i.e. 0.734) but most importantly, the smallest RMSE (i.e. 0.060 µatm·yr^{-1})

(Figure 4.4e). The slightly higher RMSE for the 2° and 1° latitude bands MLR analyses (i.e. 0.064 µatm·yr^{-1}) compared to the one on 5° suggests that the North Atlantic observational coverage (both temporal and spatial) is not sufficient to capture the area-wide multi-decadal variability. Indeed, as the area over which the MLR is trained gets narrower, the statistical model might overfit the trained dataset and therefore provide a slightly poorer predictive skill for the basin-wide region than when trained over wider areas. By developing the MLR over a different range of area width, we are therefore able to identify strengths and weaknesses in each method.

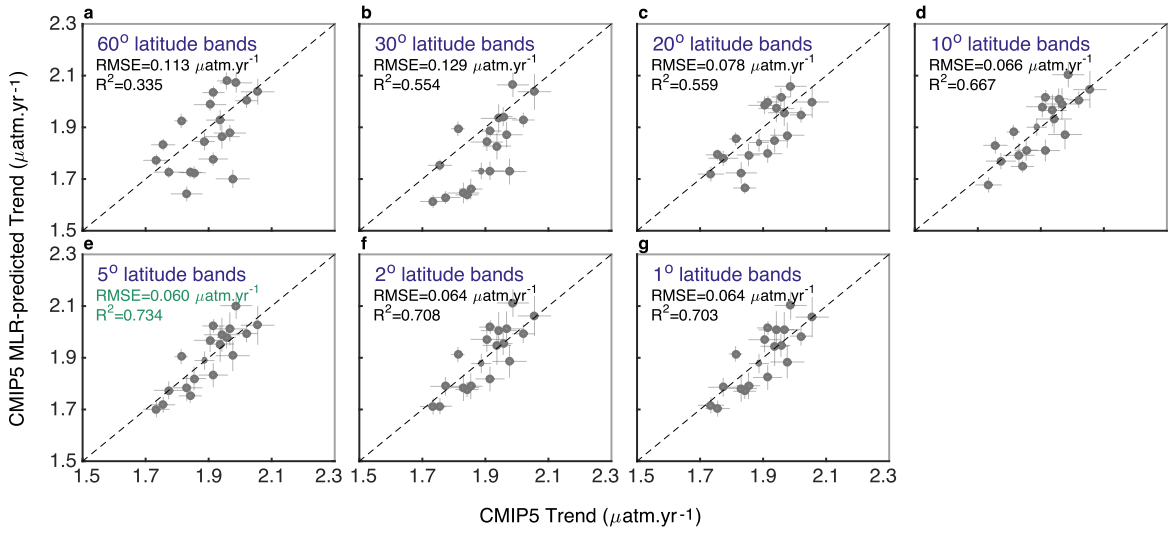

Figure 4.4: Trend uncertainty from the CMIP5-based MLR analyses. Each panel corresponds to the results from one of the seven spatially-divided MLR method. North Atlantic $fCO_{2\text{-ocean}}$ linear trends calculated over the period 1992-2014 from the CMIP5-based MLR (i.e. $\Gamma_{\text{MLR-predicted}}$) versus the CMIP5 model-truth value (i.e. $\Gamma_{\text{model-true}}$) (circles), with the corresponding standard error returned by the linear model (vertical and horizontal lines). Each panel includes the R-squared value and the RMSE between the two products. The dashed line indicates the 1-to-1 line. Points above the 1-to-1 line indicate that, for the corresponding CMIP5 models, the MLR overestimates the model-truth $fCO_{2\text{-ocean}}$ trend.

Based on the current observational coverage, the MLR analysis generated over 5° latitude bands is the optimal method for capturing the model-truth $fCO_{2\text{-ocean}}$ trend. The interpolation uncertainty on the linear trend, which will be used as the uncertainty on the observation-based $fCO_{2\text{-ocean}}$ trend (Section 4.4), is therefore calculated as the standard deviation σ of the difference between $\Gamma_{\text{MLR-predicted}}$ and $\Gamma_{\text{model-truth}}$ across the 19 CMIP5 models, based on MLR

analysis computed over latitude bands of 5° width:

$$\sigma = \sqrt{\sum_{m=1}^{19} \frac{\left(\Delta\Gamma_{\mathrm{m}} - \overline{\Delta\Gamma}\right)^2}{18}} \tag{4.7}$$

where *m* corresponds to each of the individual 19 CMIP5 models and with:

$$\Delta\Gamma_{\mathrm{m}} = \Gamma_{\mathrm{MLR-predicted,m}} - \Gamma_{\mathrm{model-true,m}} \quad \text{and} \quad \overline{\Delta\Gamma} = \frac{1}{19} \cdot \sum_{m=1}^{19} \Delta\Gamma_{\mathrm{m}} \tag{4.8}$$

Based on the above results (Sections 4.3.1 and 4.3.2) and for consistency purposes, the MLR built over 5° latitude bands across the North Atlantic is chosen as the optimal method for both reproducing the basin-wide annual means in surface $fCO_{2\text{-ocean}}$ and linear trend over the period 1992-2014. As such, the MLR-based results presented below and in Chapter 5 are drawn from the MLR analysis over 5° latitude bands only (the results from the other MLR methods are discarded).

4.4 The recent change in the North Atlantic surface $fCO_{2-ocean}$

Now that the annually-varying and trend uncertainties have been optimally determined using the CMIP5-based MLR results, the results from the observation-based MLR analysis (the one achieved on 5° latitude bands) can be explored and delimited by those uncertainties.

The North Atlantic surface $fCO_{2\text{-ocean}}$ increased approximately linearly at a rate of 1.47 ± 0.06 µatm·yr^{-1} over the period 1992-2014 (Figure 4.5a), which is much less than any of the CMIP5 models as shown by the x-axis values on Figure 4.4 (further analysis and discussion on the comparison of observation-based and CMIP5 models results will be made in Chapter 5). The 1σ uncertainty on the $fCO_{2\text{-ocean}}$ trend (i.e. 0.06 µatm·yr^{-1}) was determined through the CMIP5-based MLR analysis (Section 4.3.2; Equations 4.7, 4.8). The standard error on the trend due to the linear fit is 0.04 µatm·yr^{-1}, which is therefore encompassed by the 1σ interpolation uncertainty. The increase of surface $fCO_{2\text{-ocean}}$ is considerably less than of

$fCO_{2\text{-atmosphere}}$ over the same interval (1.88 $\pm$ 0.02 µatm·yr^{-1}), resulting in an increased atmosphere-ocean fCO_2 gradient and a potential increased ocean CO_2 uptake (implications of the $fCO_{2\text{-ocean}}$ signal on the change in the air-sea CO_2 flux are explored in Section 4.7).

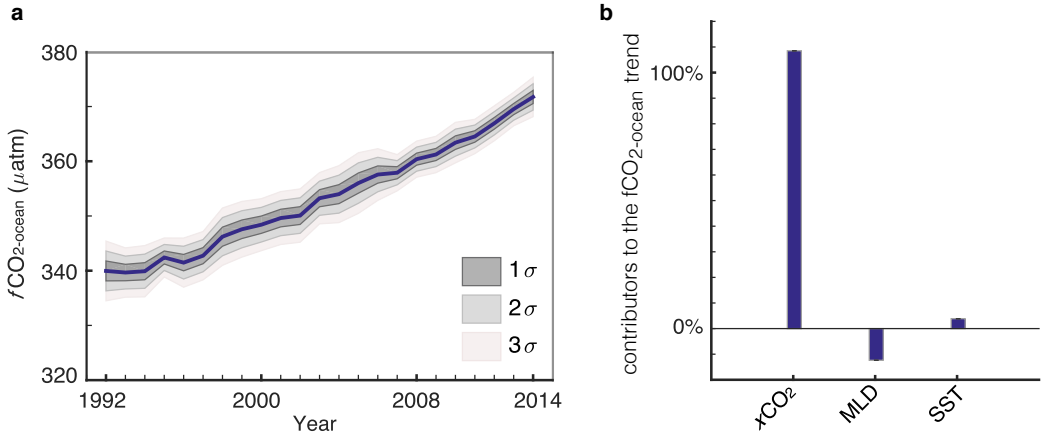

Figure 4.5: Annually averaged North Atlantic $fCO_{2-ocean}$. **a**, North Atlantic area-weighted annual averages of the surface $fCO_{2\text{-ocean}}$, taken from the observation-based MLR (blue). The dark, medium and light grey shadings represent the $\pm$ 1, 2, and 3σ annual uncertainties obtained from the CMIP5-based MLR analysis, respectively (Figure 4.3e). **b**, Contribution of the xCO_2, MLD and SST variables to the predicted $fCO_{2\text{-ocean}}$ trend (Equations 4.9, 4.10 and 4.11, respectively). All results are based from the 5° latitude band width MLR method.

To understand (to a first degree) the drivers of the trend in the North Atlantic surface $fCO_{2\text{-ocean}}$, the role of each explanatory variable in explaining the variability in the MLR-predicted $fCO_{2\text{-ocean}}$ is studied. The regression coefficients from the observation-based MLR based on latitude bands of 5° width were applied to the varying xCO_2, MLD and SST one at a time (Equations 4.9, 4.10 and 4.11, respectively). For example, to study the contribution of xCO_2 in determining the rate of change of the surface $fCO_{2\text{-ocean}}$, a new (unrealistic) $fCO_{2\text{-ocean}}$ field was calculated applying the regression coefficients to all the explanatory variables that were kept to their mean values, except for xCO_2 which was kept as the varying variable (Equation 4.9). The same approach was used for MLD and SST (Equations 4.10 and 4.11), but not for longitude as it is a non temporally varying variable.

$$fCO_{2-\text{ocean} <- xCO_2} = \beta_0 + \beta_1 \cdot \overline{\text{SST}|} + \beta_2 \cdot \overline{\text{MLD}|} + \beta_3 \cdot xCO_2| + \beta_4 \cdot \overline{\text{Lon}|} \quad (4.9)$$

$$fCO_{2-\text{ocean} <- \text{MLD}} = \beta_0 + \beta_1 \cdot \overline{\text{SST}|} + \beta_2 \cdot \text{MLD}| + \beta_3 \cdot \overline{xCO_2|} + \beta_4 \cdot \overline{\text{Lon}|} \quad (4.10)$$

$$fCO_{2-\text{ocean} <- \text{SST}} = \beta_0 + \beta_1 \cdot \text{SST}| + \beta_2 \cdot \overline{\text{MLD}|} + \beta_3 \cdot \overline{xCO_2|} + \beta_4 \cdot \overline{\text{Lon}|} \quad (4.11)$$

By studying the sign and amplitude of the trends in $fCO_{2\text{-ocean}<-xCO_2}$, $fCO_{2\text{-ocean}<-MLD}$ and $fCO_{2\text{-ocean}<-SST}$, the dominant driver (among the given explanatory variables) in controlling the rate of change in surface $fCO_{2\text{-ocean}}$ and the mechanisms involved in the surface $fCO_{2\text{-ocean}}$ can be identified. The $fCO_{2\text{-ocean}}$ trend after varying xCO_2, MLD and SST are 1.60 µatm·yr^{-1}, -0.18 µatm·yr^{-1}, and 0.057 µatm·yr^{-1}, corresponding to 108.5%, -12.4% and 3.9% of the observation-based trend (i.e. 1.47 µatm·yr^{-1}), respectively (Figure 4.5b). By having the largest trend amplitude (i.e. 1.60 µatm·yr^{-1}), the atmospheric xCO_2 seems to be the predominant driver explaining the increase in surface $fCO_{2\text{-ocean}}$ over the period 1992-2014, relative to the other explanatory variables (Figure 4.5b). Indeed, as CO_2 concentrations increase in the atmosphere due to human activity, the surface waters proportionally respond to such signal in order to reach equilibrium with the atmosphere. Over the period 1992-2014, the MLD and SST play a less important role than the atmospheric xCO_2 in controlling the surface $fCO_{2\text{-ocean}}$. The signs of the trends in $fCO_{2\text{-ocean}<-MLD}$ and $fCO_{2\text{-ocean}<-SST}$ are consistent with our understanding of the mechanisms between the explanatory variables (i.e. MLD and SST) and the surface $fCO_{2\text{-ocean}}$. Indeed, the negative sign in the $fCO_{2\text{-ocean}<-MLD}$ trend (i.e. -0.18 µatm·yr^{-1}) demonstrates that over the period 1992-2014 the change in MLD is negatively correlated with the change in surface $fCO_{2\text{-ocean}}$. For instance, an overall increase in the MLD would lead to a decrease in the surface $fCO_{2\text{-ocean}}$, suggesting that the enhancement of the dilution of CO_2 in the mixed layer and/or the stimulation of the biological activity from the input of nutrient-enriched deep waters to the surface are the main MLD-related mechanisms involved in the surface $fCO_{2\text{-ocean}}$ for the basin-wide North Atlantic (localised effect of the MLD on surface $fCO_{2\text{-ocean}}$ will be discussed in Section 4.6). As such, the impact of DIC-enriched deep waters to the surface as MLD deepens appears to be minimal over the period 1992-2014 and unlikely explains the rate of change of surface $fCO_{2\text{-ocean}}$ as MLD increases. The positive sign in the $fCO_{2\text{-ocean}<-SST}$ trend (i.e. 0.057 µatm·yr^{-1}) indicates that over the period 1992-2014 an increase in the SST leads to an increase in the surface $fCO_{2\text{-ocean}}$, which is coherent with the decrease in the solubility of CO_2 into seawater as surface waters warm. The fact that the amplitude of the trends in the the North Atlantic $fCO_{2\text{-ocean}<-MLD}$ and $fCO_{2\text{-ocean}<-SST}$ is much smaller than of the trend in the anthropogenically-driven $fCO_{2\text{-ocean}<-xCO_2}$ suggests that internal variability over the interval of study plays a minor

role [3]. While the present work enables the identification of the dominating contributors to the observed-based $fCO_{2\text{-ocean}}$ trend (i.e. the atmospheric xCO_2), the causality cannot be affirmed. Indeed, if any variable was increasing at the same rate as xCO_2 over the period 1992-2014, this variable would also be found as the contributing variable to the surface $fCO_{2\text{-ocean}}$ trend. Nevertheless, no other physical or biogeochemical variable was found to be increasing at the same rate as xCO_2, which reinforces the suggestion that the recent change in surface $fCO_{2\text{-ocean}}$ is mostly anthropogenically-driven.

4.5 Validating observation-based $fCO_{2-ocean}$ and uncertainty estimates

Although the CMIP5-based MLR analysis enables the quantification of the basin-wide uncertainty associated with the interpolation technique, to give us confidence in this approach, the behaviour of the observation-based MLR is also assessed. In this section, three different validation assessments (independent from the CMIP5-based MLR results) are carried out: a bootstrapping analysis (Section 4.5.1), the comparison with new $fCO_{2\text{-ocean}}$ data that are stored in the latest version of SOCAT, the version 5 (Section 4.5.2) and the comparison with independent interpolated products (Section 4.5.3). Finally, to justify the use of the uncertainty deduced from the CMIP5-based MLR on results from the observation-based MLR, an analysis on the subsampled $fCO_{2\text{-residual}}$ is performed (Section 4.5.4).

4.5.1 Bootstrapping analysis

To quantify the predictive skills of the interpolation technique and therefore its accuracy, a bootstrapping analysis on the observational data was undertaken. The bootstrapping analysis followed three steps: (1) excluding some of the initial data, thus defining a reduced dataset, (2) producing a MLR on latitude bands of 5° width across the North Atlantic using the reduced dataset, and (3) comparing the new MLR-predicted field with the original field at the excluded locations only.

[3] However over multi-decadal timescales, the change in MLD and SST would likely be anthropogenically dependent, as the increase of CO_2 in the atmosphere leads to an increase in global surface temperature and hence stratification.

	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	Total
R	0.72	0.73	0.73	0.72	0.71	0.68	0.69	0.76	0.77	0.78	0.73 ± 0.3
RMSE	19.2	18.9	18.9	19.3	19.6	20.3	19.4	16.4	16.3	16.1	18.4 ± 1.5
n° t-test=1	2%	4%	5%	4%	7%	2%	7%	8%	6%	3%	-

Table 4.1: Assessment of the MLR predictive skill through a set of bootstrapping analyses. Correlation R and RMSE (in µatm) in the $fCO_{2\text{-ocean}}$ between the observations and the values predicted by the MLR, for the different bootstrapping analyses. The last column indicate the mean and standard deviation across the 10 different bootstrapping analyses. Number (in %) of statistically significant bias (i.e. t-test=1) in 100 randomly generated ensembles of $fCO_{2\text{-residuals}}$, at the 5% significance level and using t-test statistics. For further details on the ensembles, see text.

Between 1992 and 2014, there are 39,831 gridded $fCO_{2\text{-ocean}}$ values (from the SOCATv4 monthly gridded product) available in the North Atlantic (excluding shelf seas). Successively and randomly 5% and 10% to 90% of the initial $fCO_{2\text{-ocean}}$ dataset were removed in 10% steps, resulting in 10 different bootstrapping analyses. To produce a coherent statistical study across the 10 different bootstrapping analyses, each statistical measure (here the correlation R and the RMSE) should be performed on an identical ensemble size. Indeed, when removing 90% of the initial observations, there are 35,848 data points from which the MLR predictive skill can be determined, while when removing 5% from the initial observational dataset, there are 1,992 data points. To create a uniform ensemble size across the bootstrapping analyses, a hundred different datasets, each of them made of a randomly-selected 100 predicted values (from the excluded locations), were successively produced for each bootstrapping analysis. For instance, for the bootstrapping generated after removing 5% of the original $fCO_{2\text{-ocean}}$ values, 100 random values among the 1,992 available for the predictive analysis were selected, from which the correlation R (between the MLR-predicted $fCO_{2\text{-ocean}}$ values and the corresponding observations) and the RMSE across the $fCO_{2\text{-residuals}}$ were calculated; this step was repeated another 99 times. A total of 100 R and RMSE values were therefore generated for each bootstrapping analysis, from which the mean correlation and RMSE values were calculated ("Total" column in Table 4.1). For a specific configuration of 100 random selections (different across the bootstrapping analyses), the correlation R between the original and predicted $fCO_{2\text{-ocean}}$ values is 0.73 ± 0.3, and the RMSE is 18.4 ± 1.5 µatm, where the standard deviation

corresponds to the spread across the 10 different bootstrapping analyses (Table 4.1). This is in agreement with a previous study (Schuster et al., 2013), which used a different MLR technique and found a RMSE of 19.9 µatm. For each bootstrapping analysis and for each sample of 100 $fCO_{2\text{-residuals}}$, a t-test was performed in order to identify potentially significant biases introduced by the MLR (leading to a total of 100 tests per bootstrapping analysis). The total number of significant biases (at the 5% significance level) is less than 10% across the bootstrapping analysis (i.e. the MLR has a lower than 10% risk to introduce a bias on a 100 residual values distribution).

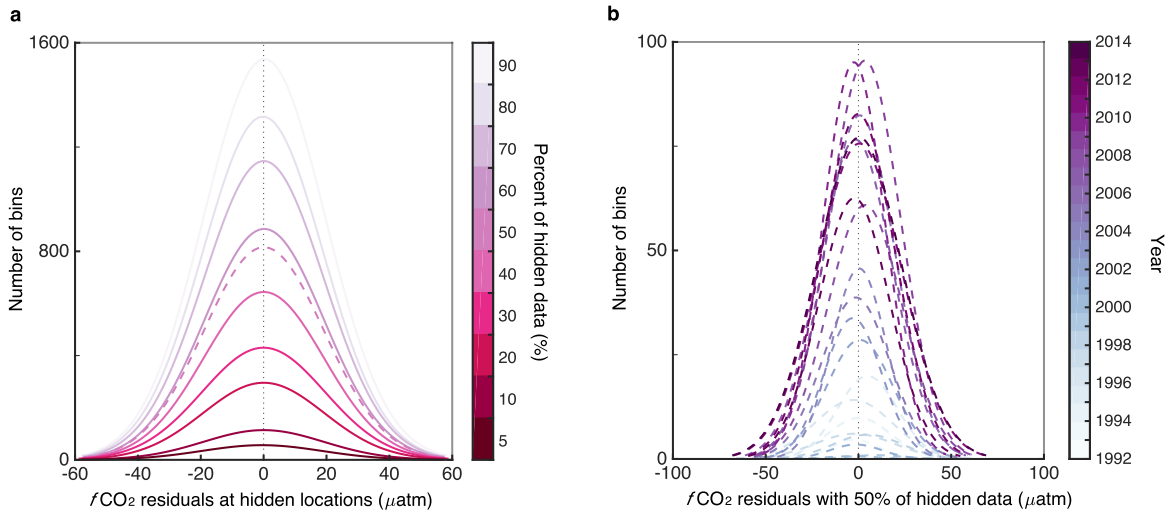

Figure 4.6: Assessment of the observation-based MLR: the bootstrapping analysis. Normal density function obtained from histograms of the $fCO_{2\text{-residuals}}$, resulting from the bootstrapping analysis and determined at the excluded locations only. **a**, each distribution presents the residuals for each of the 10 different bootstrapping analyses, in which were excluded successively 5 to 90% of the initial SOCATv4 dataset over which the MLR is trained. The dashed line highlights results when excluding 50% of the data. **b**, each distribution presents the residuals for each year, from 1992 (light) to 2014 (dark), resulting when excluding 50% of the initial SOCATv4 dataset.

When using the entire ensemble of $fCO_{2\text{-residuals}}$ defined at all the excluded locations, the distributions present no mean-bias for the bootstrapping analyses hiding from 5 to 70% of the initial data (Figure 4.6a): each distribution comes from a distribution with mean zero at the 5% significance level (t-test statistics). When hiding 80 and 90% of the initial data, the training dataset becomes too small (7,966 and 3,983 data points) and the distribution of the residuals becomes significantly biased, at the 5% significance level (t-test statistics). These results suggest that the interpolation technique, which uses all the available observations from 1992 to 2014 in the North Atlantic, does not add an overall mean bias.

To evaluate the temporal behaviour of the interpolation technique, we studied the distributions of the $fCO_{2\text{-residuals}}$ (at the excluded locations) for each year, using results from the bootstrapping analysis that removed 50% of the initial data (Figure 4.6b). After calculating the mean of each of the 23 distributions, no significant trend in the mean $fCO_{2\text{-residuals}}$ is found, at the 5% significance level. This temporal analysis was repeated on the remaining 9 bootstrapping analysis and identical results were found. Also, for each of the bootstrapping analyses, the annual area-weighted means present no significant trend in the $fCO_{2\text{-residuals}}$, at the 5% significance level. Such results, which are observation-based only, reinforce the fact that the interpolation technique is robust: the MLR does not add a bias (at least given the observation coverage) on the mean and more importantly on the trend, which is the focus of the study carried out in Chapter 5.

4.5.2 Comparison with new data

To further study the MLR's predictive skill and verify that the MLR successively performed over latitude band of 5° width is appropriate for the observational-based analysis, the MLR-predicted $fCO_{2\text{-ocean}}$ values are compared against new $fCO_{2\text{-ocean}}$ data, that are stored in the monthly gridded SOCAT product version 5 (hereinafter referred as SOCATv5; Bakker et al., 2016). The North Atlantic $fCO_{2\text{-ocean}}$ values over the period 1992-2014 that are included in SOCATv5 and do not exist in the previous version are referred as the "new observations", leading to the retrieval of 2,149 new observations.

For each MLR method (i.e. the MLR analyses computed over 60°, 30°, 20°, 10°, 5°, 2° and 1°), the differences between the MLR-predicted $fCO_{2\text{-ocean}}$ values and the new observations (i.e. the $fCO_{2\text{-residuals}}$) are studied (Figure 4.7). The RMSE of the $fCO_{2\text{-residuals}}$ (Figure 4.7h) decreases as the width of the latitude band decreases, except for the MLR built on latitude bands of 1° width which provides a RMSE slightly higher than the MLR based on 2° or 5° (i.e. 24.3 µatm against 24.0 µatm, respectively). As a reference, the RMSE of the differences between the MLR-predicted $fCO_{2\text{-ocean}}$ values and the observations stored within SOCATv4, which is based on 39,831 data points, decreases as the width of the latitude band over the MLR is built decreases (across all methods) from 24.5 µatm to 18.4 µatm for the MLR analyses built over latitude bands of 60° and 1° width, respectively. As such, while the MLR trained over latitude bands of small width (i.e. 1°) returns better $fCO_{2\text{-ocean}}$ prediction values

at the locations and times at which the measurements were made than when trained over wider latitude bands, the MLR actually becomes too specific to the trained data and therefore provides a slightly poorer predictive skill outside those measurements, which illustrate the issues involved in overfitting conditions. For the MLR trained over 5° latitude band width, the RMSE equals 24.0 µatm and 19.0 µatm for the $fCO_{2\text{-residuals}}$ using SOCAT version 5 (i.e. new data) and version 4 (i.e. training data), respectively. The difference between the two RMSE (i.e. 5 µatm) is smaller than in an independent study (Landschützer et al., 2013), which found, based on a different interpolation technique, a difference in the RMSE of $\sim$13 µatm between new and training data (i.e. using SOCAT version 1.5 and 2, respectively). This result suggests that the MLR method developed here (the one based on 5° latitude bands) captures most of the variability of the surface CO_2 system, at least based on our current understanding of the measured CO_2 system.

Across the seven different MLR methods, the interpolation technique introduces, at the locations and times of the new measurements, a negative bias (i.e. the mean $fCO_{2\text{-residuals}}$; Figure 4.7h). The bias becomes less negative as the width of the latitude bands over which the MLRs are performed decreases, except for the 1° method. Nevertheless, the bias is impacted by extreme negative values (e.g. below -75 µatm), leading to medians of the $fCO_{2\text{-residuals}}$ that are between 3 to 10 times less negative than the mean errors (Figure 4.7h). Further investigations (not shown) on the locations and times at which occurred the extreme negative $fCO_{2\text{-residuals}}$ values (i.e. arbitrarily defined as below -75 µatm) showed that these values mostly took place in May 2012 on grid cells going from around Florida to the Mediterranean Sea on an approximately horizontal line, which seems to be consistent with the path a single ship track. While high $fCO_{2\text{-residuals}}$ may be expected in localised areas of strong variability (e.g. Labrador Sea) and spread over various months due to potentially poorly represented intense variability by the linear model, the underestimation by the MLR analysis may in fact arise from a potential positive bias that would be voyage-specific. Indeed, when studying the distribution of the $fCO_{2\text{-ocean}}$ values stored in SOCATv5 in the open waters of the North Atlantic and over the period 1992-2015, the observations of May 2012 are between 2σ to beyond 3σ of the distribution: the distribution has a mean and standard deviation of 357 µatm and 29 µatm, respectively, while the minimum and maximum values of the "suspicious" data are 423 µatm and 486 µatm. As such, the investigation of the

relatively low $fCO_{2\text{-residuals}}$ values suggests that while independent observations are essential to assess the predictive skill of a linear model on localised space and time scale, one should remain cautious with the use of observational data that may contain systematic biases. Nevertheless, despite the local observations that the MLR analyses (questionably) poorly represents, the $fCO_{2\text{-residuals}}$ of each MLR method come from a distribution of mean zero, at the 5% significance level (t-test statistics). As such, at the locations and times of the new observations, the MLR does not add a significant bias (at the 5% significance level).

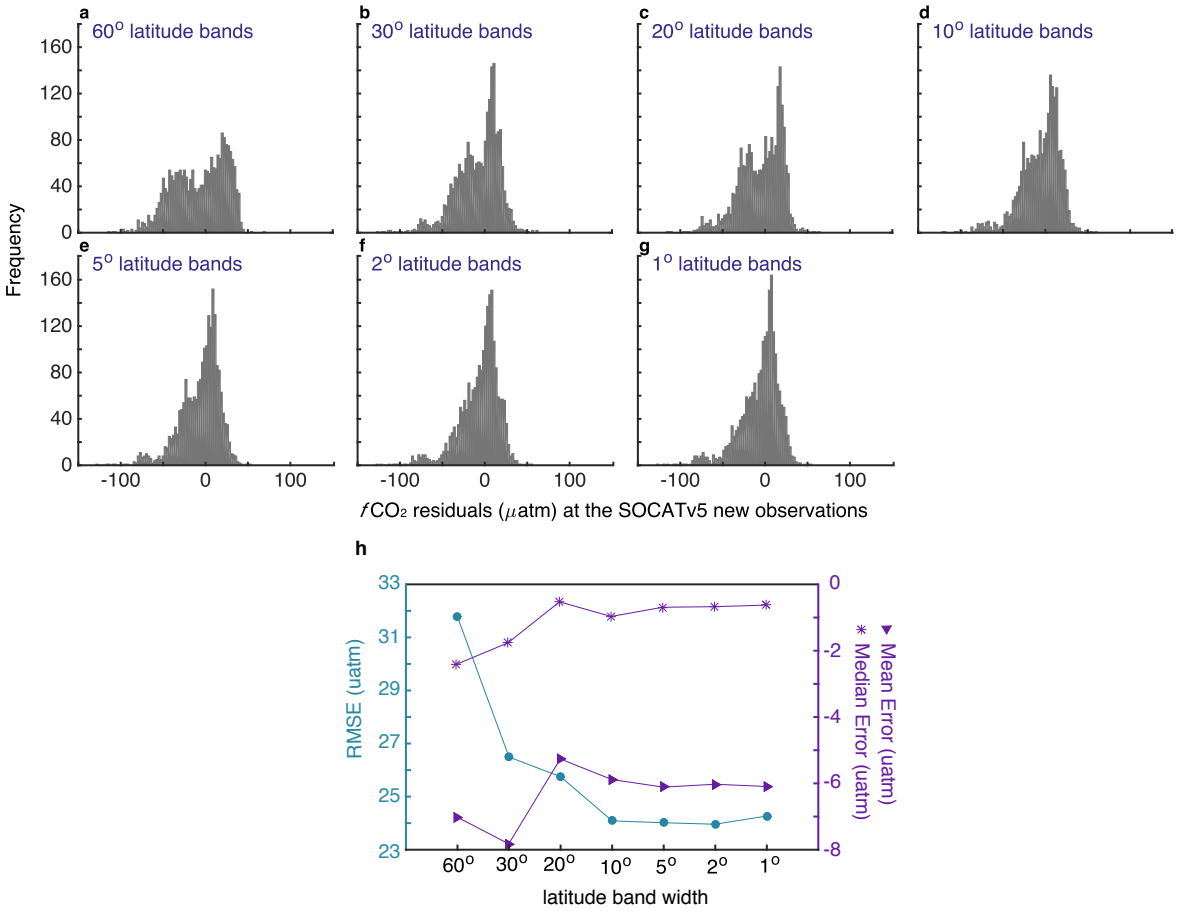

Figure 4.7: Validation using new observations. **a-g** Histogram of the difference between MLR-predicted $fCO_{2\text{-ocean}}$ by each of the seven MLR methods and the new observations stored in SOCATv5 (Bakker et al., 2016), defining the $fCO_{2\text{-residuals}}$. **h**, for each MLR method, summary over the main statistical measures of the $fCO_{2\text{-residuals}}$: RMSE in blue (circle symbol), mean and median values in purple with triangle and cross symbols, respectively.

4.5.3 Comparison with previous studies

To bring further confidence in the annual time series of the MLR-predicted $fCO_{2\text{-ocean}}$ and the associated time-varying uncertainty, the present results (Figure 4.5) are compared with three independent studies (Figure 4.8): (1) the neural network study from Landschützer et al. (2015), which provided global $fCO_{2\text{-ocean}}$ field into a $1^\circ \times 1^\circ$ grid from 1982 to 2011, (2) the linear regression analysis of Iida et al. (2015), which produced global $pCO_{2\text{-ocean}}$ into a $1^\circ \times 1^\circ$ grid from 1990 to 2016, and (3) the Jones et al. (2015) statistical gap-filling method, which generated a global $fCO_{2\text{-ocean}}$ field with a corresponding uncertainty into a $2.5^\circ \times 2.5^\circ$ grid from 1985 to 2011 (further information on the data processing of those products in Table A.5). The North Atlantic and the period 1992-2014 (if available) were extracted from the three independent products (including the uncertainty dataset) and shelf waters (i.e. waters shallower than 1,000 m depth) were removed using the ETOPO1 product (Amante and Eakins, 2015). Finally annual area-weighted averages of $pCO_{2\text{-ocean}}$, $fCO_{2\text{-ocean}}$ and uncertainty fields were calculated (Figure 4.8).

Overall, the time-varying MLR results are consistent with previous work: the time series from the three independent studies are within our time-varying and trend uncertainty estimates. Indeed, the methods from Landschützer et al. (2015), Iida et al. (2015) and Jones et al. (2015) obtained a $fCO_{2\text{-ocean}}$ (or $pCO_{2\text{-ocean}}$) trend of 1.44 ± 0.06 µatm·yr^{-1}, 1.71 ± 0.03 µatm·yr^{-1} and 1.52 ± 0.061 µatm·yr^{-1} over their corresponding intervals, respectively (where the uncertainty corresponds to the standard error returned by the linear fit), which is comparable to our trend estimate of 1.47 ± 0.06 µatm·yr^{-1}. While the results from Iida et al. (2015) are at the limit of agreement with the present work, the lack of time-varying uncertainties, which would likely be greater than the uncertainty on the linear fit (i.e. 0.03 µatm·yr^{-1}), might alter the inter-study comparisons. Also as previously discussed (Section 4.1.3), the uncertainty provided by Jones et al. (2015) covers a wide range of possible $fCO_{2\text{-ocean}}$ trends within that interval, which would lead to rather inconclusive trend studies. As such, the present work contributes to an improvement on previous uncertainty estimates, with a surface CO_2 system that is better defined.

Regarding the validation of the range in the time-varying interpolation uncertainty (i.e. 1σ, corresponding to the dark grey shading in Figure 4.8), the present estimates of 1.0 and 1.8 µatm are also in agreement with Watson et al. (2009), which found for the year 2005 an

error on mean $fCO_{2\text{-ocean}}$ between 0.8 µatm and 1.8 µatm.

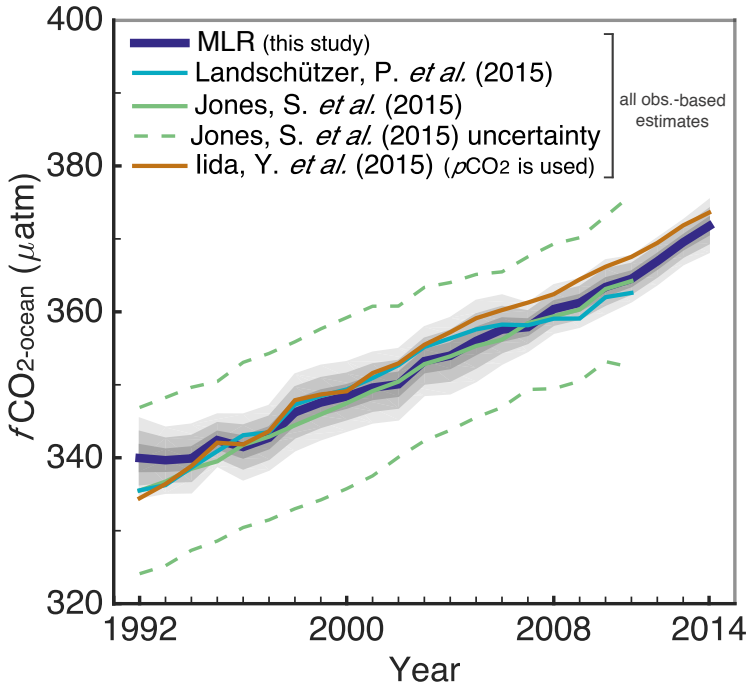

Figure 4.8: Validating the fCO_2 annual estimates with independent observation-based studies. The thick blue line presents the annually averaged $fCO_{2\text{-ocean}}$ over 1992-2014 from the observation-based MLR and the dark, medium and light grey shadings represent the $\pm$ 1, 2, and 3σ annual uncertainties obtained from the CMIP5-based MLR analysis, respectively (Figure 4.3e). The other lines present results from previous independent studies, using different techniques to reconstruct the trends, but similar data: the light blue is from the neural network method of Landschützer et al. (2015), the brown is from a multiple linear regression method from Iida et al. (2015) (this study specifically provided $pCO_{2\text{-ocean}}$, but in terms of the illustrative analysis presented here, the difference between $pCO_{2\text{-ocean}}$ and $fCO_{2\text{-ocean}}$ is not important) and the green is from the Jones et al. (2015) statistical gap-filling method, which also provided an estimate for basin-wide uncertainties shown by the dashed lines (more information on the methods used by these studies in Section 4.2).

4.5.4 Subsampled residual analysis

When applying the time-varying uncertainty determined from the CMIP5-based MLR analyses (results based on the 5° method) to the time-varying $fCO_{2\text{-ocean}}$ deduced from the observation-based MLR, one assumes that the two MLR studies react similarly to their corresponding proxy variables. For example, the model-mean $fCO_{2\text{-residuals}}$ time series is statistically not different from zero (Figure 4.3e), which does not guarantee that the basin-wide annually-varying $fCO_{2\text{-residuals}}$ from the observation-based MLR would behave in the same way as this, with a bias-free time-varying $fCO_{2\text{-residuals}}$ (which cannot be assessed).

To justify the application of the time-varying uncertainty obtained from the CMIP5-based MLR to time-varying results from the observation-based MLR, a comparative analysis on their corresponding $fCO_{2\text{-residuals}}$ is undertaken. Since the observation-based $fCO_{2\text{-residuals}}$ are only defined at places where observations were made, such comparison can only be achieved after subsampling the CMIP5-based $fCO_{2\text{-residuals}}$ at the locations, months and years of the observations. The observation-based $fCO_{2\text{-residuals}}$ was compared to the subsampled CMIP5-based $fCO_{2\text{-residuals}}$ as annually averaged area-weighted means, and more specifically to the model-mean and spread (Equations 4.4 and 4.5), respectively (corresponding respectively to the black line and grey shadings on Figure 4.9).

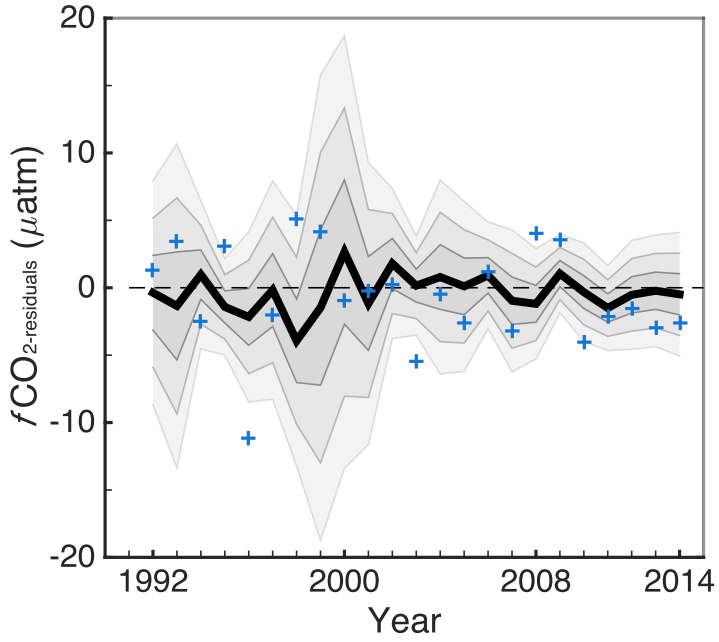

Figure 4.9: Comparison of the residuals between the CMIP5-based and observation-based MLRs. Model-mean of the annually-averaged subsampled $fCO_{2\text{-residuals}}$ (black line; Equation 4.4), with associated 1, 2 and 3 σ across the models (dark, medium and light grey, respectively; Equation 4.5). The term "subsampled" refers to the extraction of model data (here the $fCO_{2\text{-residuals}}$) at position and time where observations are available, based on SOCATv4. The blue crosses are the annual averages of the $fCO_{2\text{-residuals}}$ from the observation-based MLR analysis. Results correspond to the MLR generated over 5° latitude band width over the North Atlantic.

Overall, the patterns in the annual subsampled $fCO_{2\text{-residuals}}$ averages, from both the CMIP5-based and the observation-based MLR analyses, are impacted by the number of data points available for the different years (Figure 4.9). During the 1992-2014 period, about 70% of the observation-based annual $fCO_{2\text{-residuals}}$ (i.e. 16 years out of 23) are within 2σ

uncertainty range from the CMIP5-based MLR (i.e. 70% of the blue crosses are in the dark and medium grey shadings on Figure 4.9). Over 1992-2014, there are no significant trends in the annual $fCO_{2\text{-residuals}}$ calculated from the observation-based MLR, and from each of the subsampled CMIP5-based MLR residuals, at the 5% significance level. These results show that, over the period 1992-2014, the time-varying uncertainty calculated from the CMIP5-based MLR analysis (Figure 4.3e) is a robust estimate of the interpolation technique associated with the time-varying observation-based $fCO_{2\text{-ocean}}$.

In summary, the various validation assessments on the MLR successively built over 5° latitude bands across the North Atlantic suggest that: (1) the MLR overall provides good predictive skill that does not add a significant bias on the mean and the trend, (2) the MLR-predicted $fCO_{2\text{-ocean}}$ annual means and trend calculated over the period 1992-2014 with their associated interpolation uncertainties are in good agreement with previous independent studies, and (3) the CMIP5-based and observation-based MLR analyses behave similarly to their corresponding original $fCO_{2\text{-ocean}}$ data (at least on annual means), adding confidence to the use of the basin-wide time-varying uncertainties to delimit results from the observation-based MLR.

4.6 Discussion around the spatially distributed surface $\boldsymbol{fCO_{2-ocean}}$

Further preliminary work (not shown) indicated that the spatial subsampled residual behaviour within the CMIP5-based MLR analyses differed from the residual behaviour from the observation-based MLR (e.g. areas where the interpolation would on average overestimate the model-truth values would correspond to areas where the interpolation would on average underestimate the observed values). As such, the spatial distribution of surface $fCO_{2\text{-ocean}}$ may contain significant biases in the overall temporal mean, annual means and/or trend (the potential biases existing in the grid cells of the North Atlantic are cancelled when the basin-wide surface $fCO_{2\text{-ocean}}$ is averaged by area-weighted means; Figures 4.3 and 4.4). To fully commit to the requirements established by our MLR study (i.e. providing robust basin-wide estimates), the spatial distribution of surface $fCO_{2\text{-ocean}}$ has so far been

purposely discarded (the focus has been on area-weighted annual means, as we have been able to provide robust uncertainties for these). Nevertheless, we are aware that from a reader's point of view, the inclusion of a spatially distributed figure of the surface $fCO_{2\text{-ocean}}$ would still be a useful addition to previous independent studies, that did not provide spatially varying basin-wide uncertainty estimates (e.g. Landschützer et al., 2013). As such, the surface $fCO_{2\text{-ocean}}$ trend calculated in each grid cell of the North Atlantic for the period 1992-2014 is here presented and briefly discussed, while being aware that local biases may exist and that the spatial uncertainty may not be appropriate.

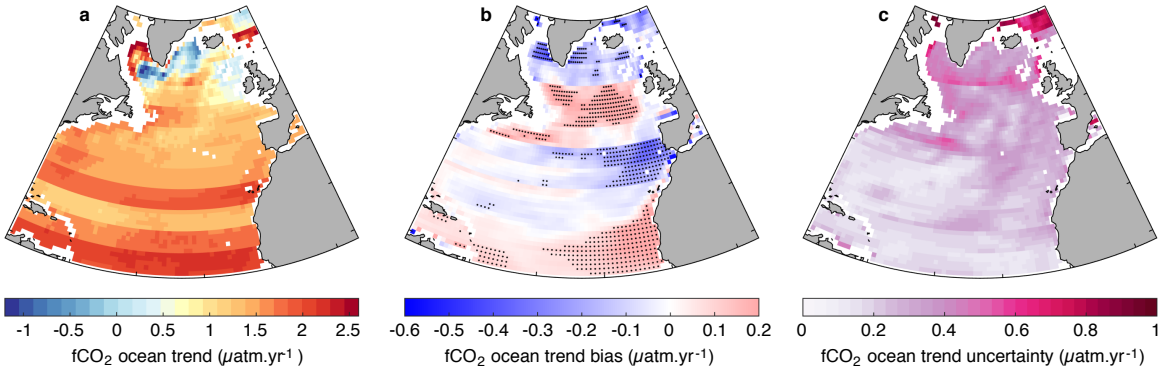

Figure 4.10: Illustrative trend in the North Atlantic $fCO_{2-ocean}$. **a**, Linear trend calculated over annual means from 1992 to 2014, using the surface $fCO_{2\text{-ocean}}$ field obtained from the observation-based MLR generated over 5° latitude band width. **b**, Bias and **c**, uncertainty on the surface $fCO_{2\text{-ocean}}$ trend, both determined using the CMIP5-based MLR analyses also generated over 5° latitude band width (see text for details). Positive biases (red in **b**) indicate that MLR interpolation technique overestimate the model-truth $fCO_{2\text{-ocean}}$ trend. Note that the results presented are purely illustrative (the application of the spatially-varying bias and uncertainty estimates taken from the CMIP5-based analyses on the observation-based spatially-varying result is not justified).

For each grid cell of the North Atlantic, the surface $fCO_{2\text{-ocean}}$ predicted by the observation-based MLR generated using 5° latitudinal band widths is averaged into annual means, which are then used to compute the surface $fCO_{2\text{-ocean}}$ linear trend over the period 1992-2014 (Figure 4.10a). The bias in the trend associated with the interpolation technique is determined in each grid cell of the North Atlantic from the CMIP5-based MLR analyses (Figure 4.10b), by specifically averaging the 19 residual trend values available in each grid cell (i.e. the residual trend is defined, for each of the 19 CMIP5-based MLR analyses, by the difference between $\Gamma_{MLR-predicted}$ and $\Gamma_{model-truth}$, where Γ is the trend in surface $fCO_{2\text{-ocean}}$ over 1992-2014; c.f. Equation 4.8). By performing a t-test in each grid cell, areas

with significant biases, at the 5% significance level, are identified (dots in Figure 4.10b). Finally, the uncertainty on the trend is also determined in each grid cell using the CMIP5-based MLR analyses, by calculating the standard deviation across the 19 residual trend values available in each grid cell (Figure 4.10c). Note that the discontinuous latitudinal changes (Figure 4.10) reflect the 5° latitudinal width division from the MLR construction.
Assuming that the spatially-varying bias and uncertainty on the $fCO_{2\text{-ocean}}$ trend determined from the CMIP5-based MLR analyses apply to the observation-based $fCO_{2\text{-ocean}}$ trend, one sees that:

- significant biases (at the 5% significance level), introduced by the interpolation technique, exist in dynamic areas of likely large variability: the Canary upwelling system, the North Atlantic Current drifting zone (i.e. the zone of transition between the subtropical and subpolar regions), the Labrador Sea, the East Greenland Current and the Gibraltar area with potential influences from the Mediterranean Sea (Figure 4.10b). As such, areas of significant biases should be considered when interpreting the spatially-varying $fCO_{2\text{-ocean}}$ trend in the real-world (Figure 4.10a). For instance, the statistically significant negative bias in the East Greenland Current area (blue dotted zone in Figure 4.10b) indicates that the MLR underestimates the model-truth trend, suggesting that the negative $fCO_{2\text{-ocean}}$ trend in that area (light blue grid cells Figure 4.10a) might in fact be less negative or even slightly positive.

- larger uncertainties are found in the subpolar region than in the subtropical region (Figure 4.10c), indicating that physical and/or biogeochemical processes occurring in the subpolar basin, critical to describe the rate of change in surface $fCO_{2\text{-ocean}}$ in that region (e.g. the influence of sea-ice formation and melting on the air-sea CO_2 flux), are not captured by the MLR. As such, the observation-based $fCO_{2\text{-ocean}}$ trends the subpolar grid cells are likely to be defined with a large uncertainty, which could prevent the detection of robust localised trend in the air-sea CO_2 flux.

- the $fCO_{2\text{-ocean}}$ trend is fairly well-homogenised across the subtropical region (i.e. no intense meridional or latitudinal gradient; Figure 4.10a). However, slightly larger $fCO_{2\text{-ocean}}$ trends are observed in the bands 10 to 20°N and 25 to 30°N relative to the rest of the subtropical region, which is potentially due to (1) the influence of enriched-DIC

intermediate waters that are brought up to the surface through the equatorial and Canary upwelling systems (Chapter 1) and that have been advected zonally via ocean circulation, and/or (2) the positive bias introduced by the MLR interpolation technique, especially in the eastern African region (Figure 4.10b). On the other hand, the $fCO_{2\text{-ocean}}$ trends present a patchy distribution in the subpolar region, especially in the Labrador Sea which displays both an intense decrease and an intense increase in surface $fCO_{2\text{-ocean}}$ over the period 1992-2014 (without bias correction; Figure 4.10a). The mechanisms involved in the surface $fCO_{2\text{-ocean}}$ trend in the subpolar are investigated below.

While the change in the North Atlantic MLD has previously been identified as being negatively correlated to the change in the North Atlantic surface $fCO_{2\text{-ocean}}$ (explained by the dilution factor and/or the stimulation of biological activity; c.f. Section 4.4), the Labrador Sea seems to present a different mechanism in explaining the rate of change in surface $fCO_{2\text{-ocean}}$. Indeed, the MLD and surface $fCO_{2\text{-ocean}}$ in the Labrador Sea present a positive correlation (R=0.77) over the period 1992-2014 (Figure 4.11a). To understand the processes occurring in the Labrador Sea, the MLD and surface $fCO_{2\text{-ocean}}$ time series are divided into three intervals:

1. the interval 1992-1996, which is characterised by an abrupt decrease in both variables. Further analyses based on the observation-based MLD product (not shown) indicated that intense deep convective events occurred in the winter months of the early[4] 1990s up until about 1996. A study based on hydrographic sections (Yashayaev et al., 2007) characterised the early 1990s as a deep convective phase for the Labrador Sea producing deep Labrador Sea Water (LSW), with a peak in the LSW formation in 1994. Indeed, the slope in the MLD becomes less negative between 1994 and 1996 compared to the years before, suggesting the influence of the deep convection event of 1994 on the basin-averaged MLD. Through deep convective events, the sinking surface watermass progressively disconnect from the atmosphere, allowing the storage of CO_2-enriched waters at depth and stimulating the air-sea CO_2 flux through the decrease in surface $fCO_{2\text{-ocean}}$ (Figure 4.11).

2. the interval 1997-2011, which is described by an approximately constant MLD and

[4]No investigation prior 1992 could be made to identify the peak of the deep convective events since the observation-based product ECCO2 (Menemenlis et al., 2008) is available from 1992.

a steadily increasing surface $fCO_{2\text{-ocean}}$. Indeed, the study of Stramma et al. (2004) identified using *in-situ* data that there was no significant formation of deep winter LSW during the years 1998-2001 (their study focused on the interval 1996-2001), but that there was shallow convections forming upper LSW during their period of study (i.e. 1996-2001); in agreement with Yashayaev et al. (2007). The absence of deep water formation in the Labrador Sea during the interval 1997-2011 (Yashayaev and Loder, 2016) therefore suggests that the surface waters of that basin were directly influenced by the change in atmospheric CO_2 concentrations and therefore responded with an approximatively linear increase in surface $fCO_{2\text{-ocean}}$.

3. the interval 2012-2014, which presents an abrupt increase in both variables. A deep convection event initiated in the winter months of 2012 and progressively intensified through the following winters, setting up the preconditioning phase for waters to subsequently sink (Yashayaev and Loder, 2016; Section 1.3.1). During this convection preconditioning phase (the phase of increasing winter MLD), progressively deeper DIC-enriched intermediate waters are brought up to the surface, leading to an increase in surface $fCO_{2\text{-ocean}}$ over the period 2011-2014.

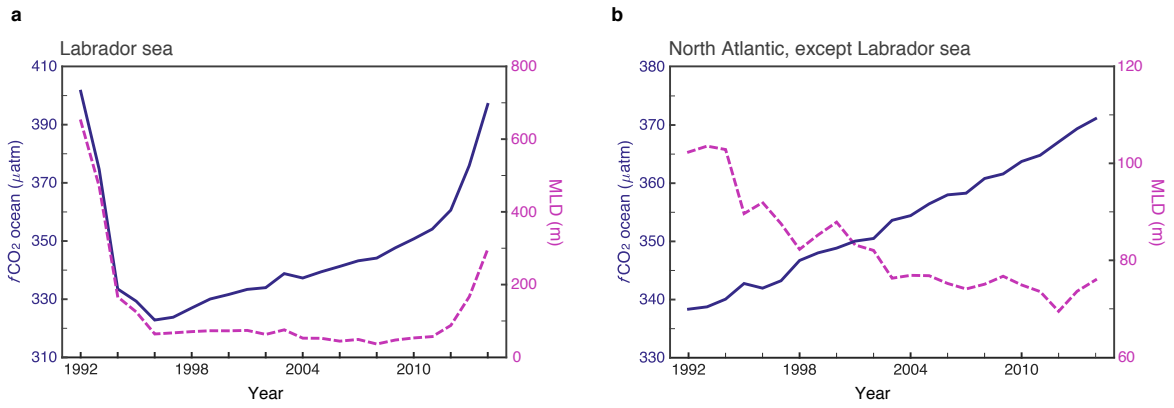

Figure 4.11: Investigating the mechanisms involved in the trend in surface $fCO_{2-ocean}$ in the Labrador Sea. Area-weighted annual time series in surface $fCO_{2\text{-ocean}}$ (blue line; left axis) and in MLD (dotted magenta line; right axis) for **a**, the Labrador Sea, and **b**, the rest of the North Atlantic. The Labrador sea is defined from 54°N to 64°N and from -62°E to -45°E.

The intense seasonal variability in the Labrador Sea, through winter convective events, explains the strong spatial variability in the surface $fCO_{2\text{-ocean}}$ trend over the period 1992-2014 in that basin; a trend that is largely influenced by changes in the MLD (as opposed to the rest

of the basin, which is driven by changes in atmospheric CO_2 concentrations; Section 4.4; Figure 4.11b). As such, the application of the trend biases and uncertainties in the Labrador Sea identified from the CMIP5-based MLR analyses would rely on the models simulating the intensity and timing of the deep water convective events[5].

4.7 Recent change in the North Atlantic air-sea CO_2 flux

With robust uncertainties on the recent change in the North Atlantic surface $fCO_{2\text{-ocean}}$ (Sections 4.3 and 4.5), the corresponding change in the air-sea CO_2 flux can be assessed, which will allow us to determine with confidence whether the North Atlantic sink recently increased or decreased. Here, the steps used to convert the $fCO_{2\text{-ocean}}$ time series into air-sea CO_2 flux time series are described (Section 4.7.1), and then results are presented (Section 4.7.2).

4.7.1 Air-sea CO_2 flux calculations

The air-sea CO_2 flux is commonly calculated at each grid cell and for each time step, before generating area-weighted annual averages and deducing the trend over the period of study (e.g. Landschützer et al., 2013). To follow the common air-sea CO_2 flux calculation steps and, more importantly, to respect the aim of this chapter as to provide robust interpolation uncertainties, each surface $fCO_{2\text{-ocean}}$ prediction held in each individual grid cell of the North Atlantic over each time step of the period 1992-2014 must be:

1. bias-free: this would be achieved if the 19 $fCO_{2\text{-residual}}$ values, held in each grid cell and provided by the CMIP5-based MLR, were to follow a distribution insignificantly different from a distribution of mean zero;

2. defined with an associated robust uncertainty: this would be achieved if, at each location and time at which observations were made, the amplitude of the $fCO_{2\text{-residual}}$ from the observation-based MLR is within the range of the CMIP5-based uncertainty (i.e. the standard deviation from the 19 $fCO_{2\text{-residual}}$ values returned by the CMIP5-based MLR).

[5]Further work (not shown) suggested that the CMIP5-models overall fail to capture the variability of the MLD in the Labrador Sea; which highlighted one of the limitations to provide spatially-varying uncertainty estimates of the $fCO_{2\text{-ocean}}$ trend or mean.

With the interpolation technique developed in this chapter, these two conditions are not satisfied, which therefore prevents us from using the spatially-interpolated $fCO_{2\text{-ocean}}$ to generate the spatial distribution of the air-sea CO_2 air-sea flux in the North Atlantic. Nevertheless, we know from the work carried out in this chapter, that the MLR-predicted $fCO_{2\text{-ocean}}$ results on annually area-weighted means are bias-free and are associated with robust uncertainties (Sections 4.3 and 4.5). As such, to respect the motivation of the present study as to provide relevant $fCO_{2\text{-ocean}}$ estimates and consequently generate a robust air-sea CO_2 flux trend estimate, the flux calculations are purely based on annual time series. While we are aware that the "time series" approach is not necessarily ideal to determine the amplitude of the North Atlantic air-sea CO_2 flux[6], we purposely focus the air-sea CO_2 flux analysis on trends, for which we are confident. We also remind that the aim of the work carried in this chapter was to (1) get fundamental new understanding of the reliability of an interpolation technique and (2) provide robust uncertainties on $fCO_{2\text{-ocean}}$ estimates, and as such, the air-sea CO_2 flux estimates mostly allow us to illustrate the implications of the $fCO_{2\text{-ocean}}$ trend and its associated uncertainty on the North Atlantic CO_2 uptake.

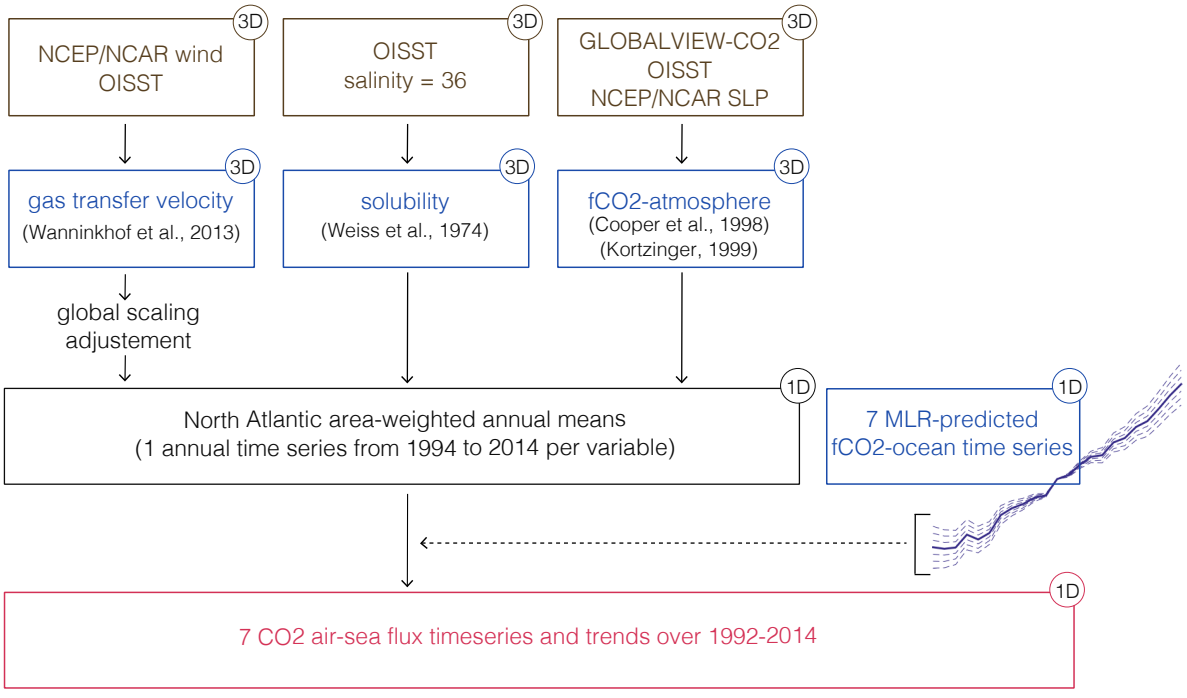

Figure 4.12: Schematic representation of the air-sea CO_2 flux calculation steps.

[6]For example, the amplitude is impacted by localised winds and would return a different value than when using calculating the flux from annually averaged K.

To generate the trend in the North Atlantic air-sea CO_2 flux, defined with an uncertainty derived from the interpolation error, the annual time series for each of the flux's components (i.e. the gas transfer velocity *K*, solubility α, $fCO_{2\text{-atmosphere}}$ and the MLR-predicted $fCO_{2\text{-ocean}}$, as used in Equation 1.4)) are produced as follow (Figure 4.12):

- The gas transfer velocity (K): *K* was determined following the definition of Wanninkhof (2014):

$$K = 0.251 \cdot \overline{W^2} \cdot \left(\frac{Sc}{600}\right)^{-0.5} \tag{4.12}$$

where *K* is in cm·h^{-1}, *W* the wind speed in m·s^{-1} and *Sc* the Schmidt number (Wanninkhof, 2014):

$$Sc = 2116.8 - 136.25 \times T + 4.7353 \times T^2 - 0.092307 \times T^3 + 0.0007555 \times T^4 \tag{4.13}$$

For the temperature *T* (used in the Schmidt number, in °C), the monthly OISST product for 1992-2014 was used (Reynolds et al., 2007) (Table A.5). For the wind speed *W*, the 6-hourly zonal and meridional winds at 10 meters during 1992-2014 were taken from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis 1 product (Kalnay et al., 1996) (Table A.5). The wind speed squared (i.e. W^2) was calculated at each 6-hour time step, before being averaged into monthly values. The shelf waters (i.e. above 1,000 m depth) of the global monthly SST and W^2 products were removed using the bathymetry from ETOPO1 (Amante and Eakins, 2015). From this, the gas transfer velocity was calculated (Equation 4.12) for each grid cell of the globe across the period 1992-2014. The global ocean average of *K* for 1992-2014 is 17.2 cm·h^{-1}. The gas transfer velocity *K* was adjusted by a scaling factor ($\sim$0.93) to reach the global mean of K$\sim$16 cm·h^{-1} to match a recent estimate with a different wind product (Wanninkhof et al., 2013; Landschützer et al., 2016). From the adjusted global *K* field, the North Atlantic was extracted and annual area-weighted means were calculated.

- The solubility (α): α was determined using the definition of Weiss (1974):

$$\alpha = exp\Bigg(-60.2409 + \frac{9345.17}{T} + 23.3585 \cdot log\Big(\frac{T}{100}\Big) + S \cdot \Bigg(0.023517 - 0.00023656T + 0.0047036 \cdot \Big(\frac{T}{100}\Big)^2\Bigg)\Bigg) \tag{4.14}$$

  where the solubility α is in µmol·kg^{-1}·µatm^{-1}, *T* is the SST in °K and *S* is the salinity. As in Watson et al. (2009), the salinity was defined as 36. For the temperature field, the monthly OISST product from 1992-2014 was used (Reynolds et al., 2007). Once the shelves waters removed from the solubility removed (Amante and Eakins, 2015), area-weighted annual means from 1992 to 2014 for the North Atlantic were generated.

- The $fCO_{2\text{-atmosphere}}$: Two steps were involved in the calculation of the 3-dimensional $fCO_{2\text{-atmosphere}}$. First, the atmospheric xCO_2 mixing ratio, taken from the reference matrix of GLOBALVIEW-CO2, was converted in the atmospheric partial pressure of CO_2 using Equations 1.2 and 1.3, for which the SST and the sea level pressure (SLP) fields were required. Second, the $pCO_{2\text{-atmosphere}}$ was converted into $fCO_{2\text{-atmosphere}}$ using Equation 1.1, which also required the SST and Sea Level Pressure (SLP) fields. For the SST field, the OISST product was used (Reynolds et al., 2007), and for the SLP field, the NCEP/NCAR product was used (Table A.5). A mask was applied to the grid cells corresponding to shelves areas (Amante and Eakins, 2015), hence providing a 3-dimensional $fCO_{2\text{-atmosphere}}$ only available at the open-waters North Atlantic grid cells, which was then spatially averaged (area-weighted) into annual means.

- The $fCO_{2\text{-ocean}}$: In order to provide an air-sea CO_2 flux that captures the influence of the interpolation uncertainty, seven $fCO_{2\text{-ocean}}$ time series were reconstructed from the MLR study. The first time series corresponds to the $fCO_{2\text{-ocean}}$ returned by observation-based MLR, referred as the mean $fCO_{2\text{-ocean}}$ time series (i.e. blue line in Figure 4.5). The second and third time series correspond to mean $fCO_{2\text{-ocean}}$ time series (which has a trend of 1.47 µatm·yr^{-1}) with a change in the trend of $\pm$ 1σ from the trend uncertainty (where σ = 0.06 µatm). As such the two generated time series have a trend of 1.46 µatm·yr^{-1} and 1.53 µatm·yr^{-1}. Similarly, the last four time series were similarly reconstructed using the $\pm$ 2 and 3σ values. The six time series that capture

the uncertainty on the trend were centred in 2003, which is the year at which the mean $fCO_{2\text{-ocean}}$ over the period 1992-2014 is reached.

In summary, one annual time series was generated for the North Atlantic across the period 1992-2014 for the gas transfer velocity, solubility and $fCO_{2\text{-atmosphere}}$, and seven for the $fCO_{2\text{-ocean}}$, leading to the production of seven air-sea CO_2 flux trends (Figure 4.12)

4.7.2 Air-sea CO_2 flux results

Over the period 1992-2014, the North Atlantic CO_2 sink increased at a rate of[7] 0.081 $\pm$ 0.012 PgC·yr^{-1}·decade^{-1}. Our result is in agreement with the air-sea CO_2 flux product generated by Landschützer et al., 2015, which estimated a trend in the flux for the North Atlantic open waters (i.e. waters shallower than 1,000 m depth were removed using ETOPO1; Amante and Eakins, 2015) of 0.069 $\pm$ 0.014 PgC·yr^{-1}·decade^{-1} over the period 1992-2011, where the uncertainty corresponds to the standard error returned by the linear fit. Our result suggests that the North Atlantic had an additional uptake of about 2.2 PgC over the 23-year interval, relative to the flux in 1992 (Figure 4.13), corresponding to about 30% of the global oceanic CO_2 sink over that interval (i.e. 1992-2014) (Le Quéré et al., 2015). Indeed, the global oceanic CO_2 sink estimated by the Global Carbon Budget is 7.9 $\pm$ 1.4 PgC over the period 1992-2014, relative to the flux in 1992 (Le Quéré et al., 2015).

To study the contribution of the $\Delta fCO_{2\text{-ocean}}$ (i.e. $fCO_{2\text{-atmosphere}}$ - $fCO_{2\text{-ocean}}$), solubility and gas transfer velocity on the air-sea CO_2 flux trend, each variable was successively varied within Equation 1.4, while the other two variables were hold at their mean value. Trends in the air-sea CO_2 flux for successively varying $\Delta fCO_{2\text{-ocean}}$, solubility and gas transfer velocity are 0.077, -1.7$\times 10^{-3}$ and 5.2 $\times 10^{-3}$ PgC.yr^{-1}.decade^{-1}, respectively, corresponding to about 95%, -2% and 6% of the mean observed trend, respectively. As such, the increase in the North Atlantic CO_2 sink is very likely to be dominated by the anthropogenic perturbation to the CO_2 system.

[7]The mean value (i.e. 0.081 PgC·yr^{-1}·decade^{-1}) corresponds to the air-sea CO_2 flux trend value calculated from the mean $fCO_{2\text{-ocean}}$ time series (i.e. blue line in Figure 4.5), and the uncertainty value (i.e. 0.012 PgC·yr^{-1}·decade^{-1}) corresponds to the absolute trend difference between the one calculated from mean $fCO_{2\text{-ocean}}$ time series (which had a trend of 1.47 µatm; blue line in Figure 4.5) and the one calculated from the $fCO_{2\text{-ocean}}$ time series that had a change in the trend of 1σ (hence with a trend of 1.47 + 1 σ µatm, where the trend uncertainty σ equals 0.06 µatm) (Section 4.7.1).

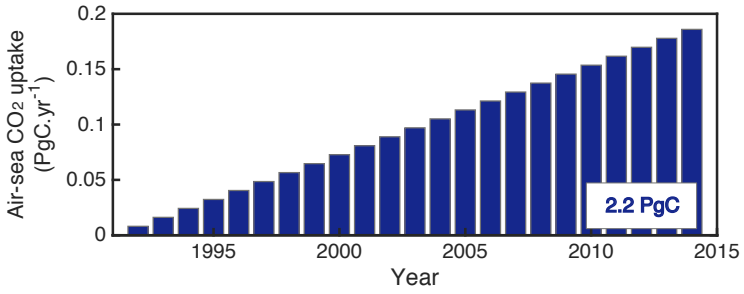

Figure 4.13: The North Atlantic CO_2 sink. The sum of the blue area (i.e. 2.2 PgC) corresponds to the North Atlantic CO_2 sink over the period 1992-2014, relative to the flux in 1992, calculated from the observation-based air-sea CO_2 flux linear trend in the North Atlantic over the period 1992-2014 (i.e. 0.0081 PgC·yr^{-2}).

4.8 Summary

To overcome the challenges associated with irregular spatiotemporal coverage in surface CO_2 measurements (Chapter 3), various interpolation methods have been developed by the community, each of them with different strengths and limitations. Until now, the quantification of the basin-wide uncertainty associated with such gap-filling methods has received little attention, limiting a thorough understanding of the recent change in the North Atlantic CO_2 sink. Here, a Multiple Linear Regression (MLR) method using both observation and model data was developed to determine the recent change in the North Atlantic surface $fCO_{2\text{-ocean}}$ with robustly quantified basin-wide interpolation uncertainty. Specifically, the MLR was successively trained over latitude bands of 5° width across the North Atlantic using SST, MLD, atmospheric xCO_2 and geographical position as proxy variables to predict the basin-wide surface $fCO_{2\text{-ocean}}$ from 1992 to 2014. By repeating the MLR analysis across 19 Earth System Models (ESMs), we were able to reproduce with confidence the annually-varying $fCO_{2\text{-ocean}}$ and trend over the period of study. From this, we show that the North Atlantic CO_2 uptake increased at a rate of 0.081 ± 0.012 PgC·yr^{-1}·decade^{-1} during the period 1992 to 2014, corresponding to an additional uptake of $\sim$2.2 PgC over this interval relative to 1992, a signal that most likely responds to the increase in CO_2 concentrations in the atmosphere due to anthropogenic activities. With these estimates of uncertainty, it is now possible to determine whether ESMs projections' underlying CO_2 emissions targets have the potential to capture the observation-based trends; an assessment that will be tackled in the following chapter (Chapter 5).

Chapter 5

Identifying and explaining model biases in the North Atlantic CO_2 sink

This chapter aims to understand the recent change in the North Atlantic surface $fCO_{2\text{-ocean}}$ in the CMIP5 models by (1) evaluating their simulated change in surface $fCO_{2\text{-ocean}}$ against well-constrained observation-based estimates, and (2) investigating the reasons behind the discrepancy between the CMIP5 outputs and the observation-based results using a set of model simulations initialised with observations.

To determine the maximum CO_2 emissions compatible with a given future climate scenario, the scientific community requires prediction of the proportion of human-emitted CO_2 that will be taken up by the land and ocean carbon reservoirs (Ciais et al., 2013). Earth System Models (ESMs) are the only way to make these critical future carbon cycle projections, yet validation of time-varying CO_2 uptake within these models remains a key challenge for two main reasons: (1) time-varying estimates of real-world ocean CO_2 concentrations have so far been lacking robust uncertainties or have been provided with wide uncertainties (Jones et al., 2015) limiting how they can be used to constrain models, and (2) models may be in a different phase of variability relative to the real-world, which would add difficulties when interpreting the simulated trends against the trends within the real-world, especially on inter-annual and decadal timescales during which the emerging signal is most likely not detectable (McKinley et al., 2011).

Here, by making use of the uncertainty assessment carried out in Chapter 4, the decadal time-varying surface CO_2 in the North Atlantic simulated by the current generation of ESMs

is robustly evaluated. Specifically, the change in the North Atlantic surface $fCO_{2\text{-ocean}}$ and air-sea CO_2 flux over the period 1992-2014 is evaluated in 19 CMIP5 models (Section 5.1). The evaluation of these 19 CMIP5 models raises two main questions: (1) Are the contradicting results between the simulated trends in the CMIP5 models and observation-based estimates resulting from differences in the phase of variability captured by the two systems (i.e. the modelled and the real-world system)? (2) If internal variability is not likely to be the main cause for the model discrepancy with the observation-based estimates, what are the main reasons explaining the bias in the CMIP5 models? By using the pre-industrial control runs provided from the available CMIP5 models and generating a set of ocean simulations initialised with observations, these two questions will be investigated in Sections 5.2 and 5.3, respectively.

This chapter will therefore bring into question our ability to predict the decadal-to-centennial evolution of the ocean sink for CO_2 in the future, in the basin where the most intense anthropogenic CO_2 sink on the planet presently exists (Khatiwala et al., 2013).

5.1 Evaluating the recent change in the North Atlantic CO_2 sink in the CMIP5 models

Through the quantification of the interpolation uncertainty on the North Atlantic surface $fCO_{2\text{-ocean}}$ carried out in Chapter 4, the recent change in the surface $fCO_{2\text{-ocean}}$ and on the air-sea CO_2 flux simulated by 19 CMIP5 models (Table A.3) are here evaluated. Prior to the model evaluation, the surface $fCO_{2\text{-ocean}}$ and air-sea CO_2 flux from the CMIP5 model outputs were downloaded and processed (e.g. conversion from $pCO_{2\text{-ocean}}$ into $fCO_{2\text{-ocean}}$ and regridding) following the method described in Appendix A.1. Additionally, to provide a consistent model-observation comparison, the models' fields (i.e. $fCO_{2\text{-ocean}}$ and air-sea CO_2 flux) for the North Atlantic[1] over the period 1992-2014 were extracted and had their shelf waters (i.e. below 1,000 m depth) removed using the ETOPO1 bathymetry product (Amante and Eakins, 2015).

We find: (1) the CMIP5 models overall behave similarly to each other in response to the

[1]The North Atlantic is defined from 10°N to 70°N and from -75°E to 5°E.

increase in atmospheric CO_2 concentrations (i.e. most of the models are clustered together), and (2) the CMIP5 models differ from the observation-based estimates on the amplitude of the surface $fCO_{2\text{-ocean}}$, and importantly its trend (Figure 5.1a, b). Over the period 1992-2014, the North Atlantic $fCO_{2\text{-ocean}}$ trend in the CMIP5 models is on average 1.90 $\pm$ 0.09 µatm·yr^{-1} (where the $\pm$ 1σ value corresponds to the inter-model variability), and ranges from 1.73 $\pm$ 0.04 µatm·yr^{-1} within the GFDL-ESM2G model to 2.06 $\pm$ 0.05 µatm·yr^{-1} within the HadGEM2-CC model (where the uncertainty corresponds to the standard error of the linear fit applied to the annual time series) (Table 5.1). This implies that the models' surface ocean concentration is closely following the rise of atmospheric CO_2 (i.e. the $fCO_{2\text{-atmosphere}}$ trend over the period 1992-2014 is 1.88 $\pm$ 0.02 µatm·yr^{-1}), consequently limiting the air-sea CO_2 gradient and therefore the removal of CO_2 from the atmosphere. Unexpectedly, the $fCO_{2\text{-ocean}}$ trends in the CMIP5 models are significantly larger than the mean of observation-based $fCO_{2\text{-ocean}}$ trend (i.e. 1.47 µatm·yr^{-1}), at the 5% significance level (right-tailed t-test statistics).

When considering an identical increase in $fCO_{2\text{-atmosphere}}$ across all the CMIP5 models and in the real-world, the fact that the CMIP5 models have a larger $fCO_{2\text{-ocean}}$ trend than in the real-world means that the difference between $fCO_{2\text{-atmosphere}}$ and $fCO_{2\text{-ocean}}$ (i.e. ΔfCO_2) increases in the models at a slower rate than in the real-world (Figure 5.1c). As such, the air-sea CO_2 flux, which is proportional to ΔfCO_2 (Equation 1.4), would increase at a slower rate in the models than in the real-world, in the absence of significant trend in the gas transfer velocity and solubility (Figure 5.1c). Indeed, over the period 1992-2014, the CMIP5 models have a mean increasing North Atlantic CO_2 uptake of 0.026 $\pm$ 0.014 PgC·yr^{-1}·decade^{-1} (Table 5.1), which is significantly smaller than the observation-based estimate (right-tailed t-test statistics at the 5% significance level) (Figure 5.1d). While the North Atlantic had a mean additional uptake of about 2.2 PgC over the 23-year interval relative to the flux in 1992 (Figure 4.13), the CMIP5 models provide on average a corresponding additional uptake of 0.72 $\pm$ 0.40 PgC, which is about 3 times smaller than the observation-based estimate.

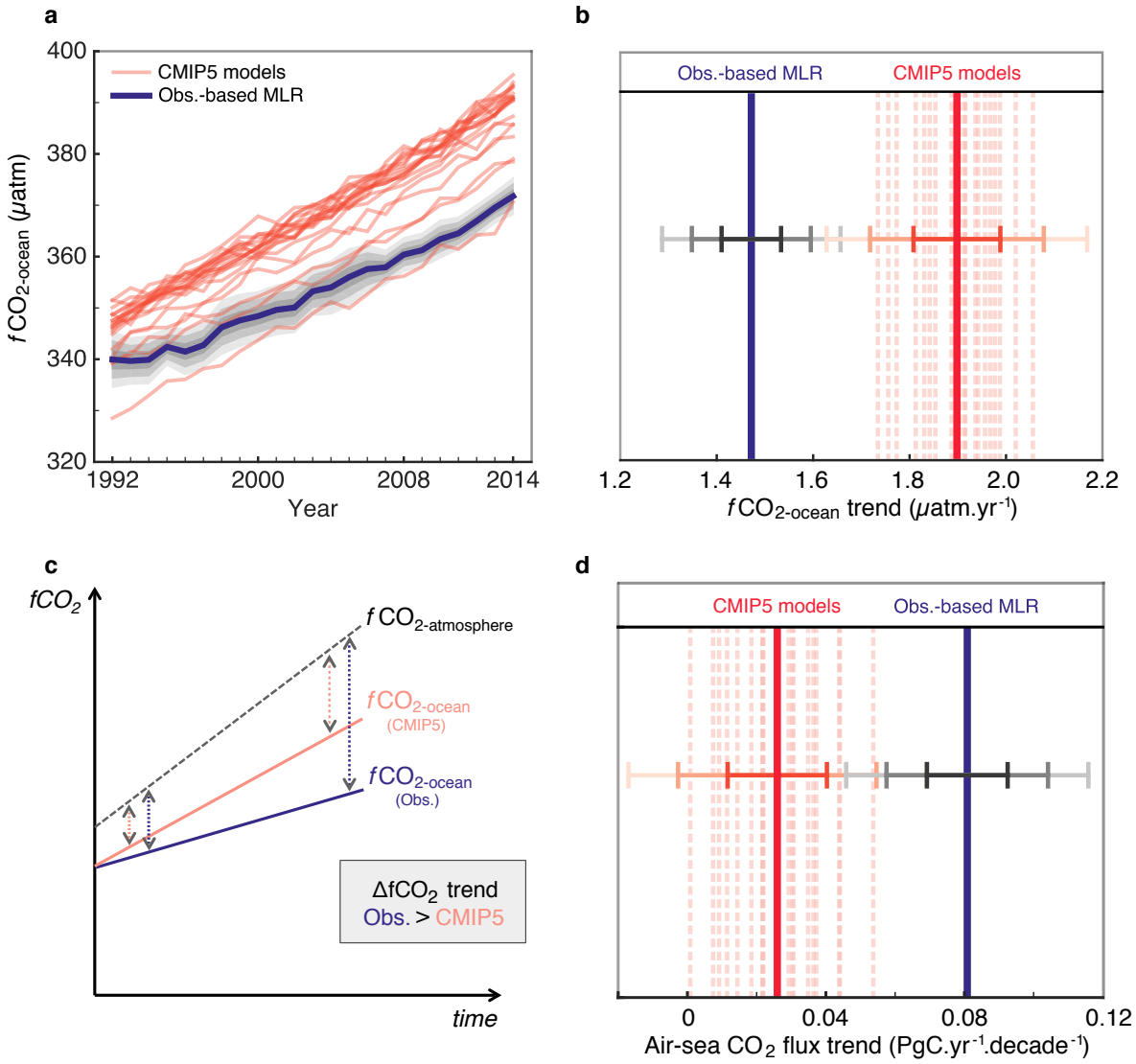

Figure 5.1: Comparison of the annually-varying and trend in the surface $fCO_{2-ocean}$ in the CMIP5 models against observation-based estimates. **a**, Annually area-weighted averaged North Atlantic $fCO_{2\text{-ocean}}$ over the period 1992-2014 and **b**, resulting linear trends. **c**, Schematic representation of the impact of the trends in $fCO_{2\text{-ocean}}$ on the trends in ΔfCO_2 (i.e. $fCO_{2\text{-atmosphere}}$-$fCO_{2\text{-ocean}}$, to which the air-sea CO_2 flux is proportional). **d**, Air-sea CO_2 flux trends calculated over the North Atlantic and over the period 1992-2014 (following the calculations of Section 4.7.1 for the observation-based estimate). In all panels, blue corresponds to observation-based MLR results (Chapter 4) and orange to each of the available CMIP5 models (Table 5.1). The dark, medium and light grey shadings (in **a**, **b**, **d**) respectively represent the $\pm$ 1, 2, and 3σ annual and trend uncertainties obtained Chapter 4 Section 4.3. In **b** and **d**, the thick orange line corresponds to the CMIP5 model-mean trend value, and the $\pm$ 1, 2 and 3σ across the models by the associated error bars.

Model Name	$fCO_{2-ocean}$ trend ($\mu atm \cdot yr^{-1}$)	Air-sea CO_2 flux trend ($PgC \cdot yr^{-1} \cdot decade^{-1}$)
CESM1-BGC	1.92 ± 0.03	$\mathbf{0.030 \pm 0.005}$
CanESM2	2.02 ± 0.03	0.0074 ± 0.0085
GFDL-ESM2G	1.73 ± 0.04	$\mathbf{0.037 \pm 0.007}$
GFDL-ESM2M	1.77 ± 0.05	$\mathbf{0.035 \pm 0.008}$
GISS-E2-H-CC	1.76 ± 0.04	$\mathbf{0.022 \pm 0.001}$
GISS-E2-R-CC	1.84 ± 0.04	$\mathbf{0.036 \pm 0.006}$
HadGEM2-CC	2.06 ± 0.05	0.014 ± 0.012
HadGEM2-ES	1.99 ± 0.05	0.00082 ± 0.0092
IPSL-CM5A-LR	1.94 ± 0.04	$\mathbf{0.029 \pm 0.011}$
IPSL-CM5A-MR	1.83 ± 0.05	$\mathbf{0.044 \pm 0.009}$
IPSL-CM5B-LR	1.94 ± 0.06	0.011 ± 0.001
MIROC-ESM-CHEM	1.98 ± 0.05	0.0092 ± 0.0090
MIROC-ESM	1.97 ± 0.05	$\mathbf{0.022 \pm 0.001}$
MPI-ESM-LR	1.96 ± 0.03	$\mathbf{0.019 \pm 0.007}$
MPI-ESM-MR	1.90 ± 0.05	0.022 ± 0.011
MRI-ESM1	1.91 ± 0.05	-
NorESM-ME	1.81 ± 0.02	$\mathbf{0.054 \pm 0.006}$
bcc-csm1-1-m	1.85 ± 0.04	$\mathbf{0.044 \pm 0.006}$
bcc-csm1-1	1.89 ± 0.03	$\mathbf{0.031 \pm 0.005}$
Model-mean	$\mathbf{1.90 \pm 0.09}$	$\mathbf{0.026 \pm 0.014}$
Observation-based	$\mathbf{1.47 \pm 0.06}$	$\mathbf{0.081 \pm 0.012}$

Table 5.1: Trends in surface $fCO_{2-ocean}$ and air-sea CO_2 flux for each CMIP5 model and from the observation-based estimates. All trends were calculated over the North Atlantic and over the period 1992-2014. Significant air-sea CO_2 flux trends are in bold, at the 5% significance level. The uncertainty on the CMIP5 models estimates corresponds to the standard error of the linear fit, except for the model-mean values which correspond to the inter-model variability. Details about the observation-based trend estimate and associated uncertainty are developed in Chapter 4.

5.2 Investigating the robustness of the model evaluation

The model evaluation on the North Atlantic surface $fCO_{2\text{-ocean}}$ trend suggested that the CMIP5 models largely underestimate the increase in North Atlantic CO_2 oceanic sink observed over the period 1992-2014 (Section 5.1). The discrepancy in the surface $fCO_{2\text{-ocean}}$ trends (and therefore in the air-sea CO_2 flux trends) between the CMIP5 models and the observation-based estimates could result from four main arguments: (1) the CMIP5 models and the real-world could be forced with slightly different atmospheric CO_2 concentrations, which would have impacted the surface $fCO_{2\text{-ocean}}$ and led to slightly different trends between the two systems (i.e. the models and the real-world), (2) the trend uncertainty on the $fCO_{2\text{-ocean}}$ observation-based estimates, which was calculated from a study across 19 CMIP5-based

MLR analyses, could in fact be sensitive to the number and/or combination of the chosen CMIP5-based MLR analyses, (3) the real-world could be experiencing, over the period of study, a phase of natural variability not captured by any of the CMIP5 models, making the two systems not directly comparable, and/or (4) the CMIP5 models could poorly represent or miss some key characteristics of the marine CO_2 system necessary to the capture of the surface $fCO_{2\text{-ocean}}$ trend. The above arguments (1), (2), and (3) are here discussed in Sections 5.2.1, 5.2.2 and 5.2.3, respectively. The investigation of the mechanisms controlling the surface $fCO_{2\text{-ocean}}$ trend in the CMIP5 models, mentioned by argument (4), will be investigated through a set of ocean simulations later in the chapter (Section 5.3).

5.2.1 A model bias due to different atmospheric xCO_2?

As discussed in the previous chapter, the observation-based $fCO_{2\text{-ocean}}$ trend over the period 1992-2014 is very likely to be dominated by the signal in the atmospheric xCO_2 (c.f. Figure 4.5). To understand the mechanisms involved in the North Atlantic CO_2 uptake in the CMIP5 models, and therefore assess whether the models' $fCO_{2\text{-ocean}}$ trends respond to similar drivers as in the real-word, the drivers of the modelled trends are determined using the same approach as the one used for the observational data. By applying the regression coefficients returned by the CMIP5-based MLR analyses to the successively varying explanatory variables (c.f Section 4.4), the xCO_2, MLD and SST time-varying signals are identified to constitute $88 \pm 3\%$, $0.4 \pm 0.9\%$ and $12 \pm 3\%$ of the models' $fCO_{2\text{-ocean}}$ trends, respectively, where the uncertainty corresponds to the the inter CMIP5-based MLR analyses. As such, in both the real-world and the model-world, the change in surface $fCO_{2\text{-ocean}}$ is highly influenced by the change atmospheric xCO_2. To verify that the discrepancy in the surface $fCO_{2\text{-ocean}}$ trends between the observation-based results and the CMIP5 models does not arise from differences in the respective atmospheric xCO_2 products, the annual atmospheric xCO_2 values used in the real-world and in the model-world are compared.

The CMIP5 models were forced using an annual and global xCO_2 time series (i.e. a single xCO_2 value was used for each grid cell of the same year) using historical measurements until 2005 and using xCO_2 projections from 2006 taken from the business-as-usual scenario (i.e. RCP8.5; Riahi et al., 2007; Taylor et al., 2012). On the other hand, the observation-based $fCO_{2\text{-ocean}}$ estimates and the associated air-sea CO_2 flux used the

GLOBALVIEW-CO_2 reference matrix (GLOBALVIEW-CO2, 2013), which is a latitudinally and monthly varying product. The GLOBALVIEW-CO_2 reference matrix was therefore extracted for the North Atlantic and area-weighted into annual means.

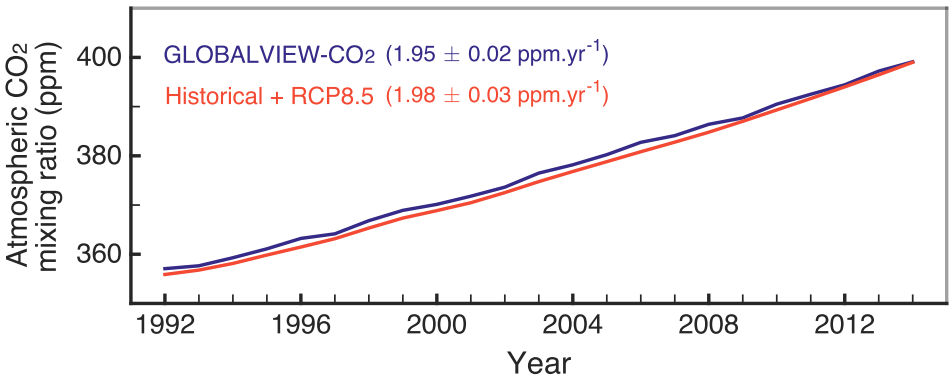

Figure 5.2: Annually-varying atmospheric xCO_2. The GLOBALVIEW-CO_2 (GLOBALVIEW-CO2, 2013) and the "historical + RCP8.5" (Riahi et al., 2007; Taylor et al., 2012) time series refer to the atmospheric xCO_2 values used to calculate the observational-based estimates and to force the CMIP5 models, respectively. The trends are calculated over the period 1992-2014, and where the uncertainty corresponds to the standard error returned by the linear fit.

Overall, the xCO_2 annual time series from GLOBALVIEW-CO_2 and from the "historical + RCP8.5" scenarios are similar, with indistinguishable resulting trends (i.e. 1.95 ± 0.02 ppm·yr^{-1} and 1.98 ± 0.02 ppm·yr^{-1}, respectively; Figure 5.2). Nevertheless, the slightly higher mean xCO_2 trend in the "historical + RCP8.5" scenarios compared to the mean GLOBALVIEW-CO_2 trend (by 0.03 ppm·yr^{-1}) might have had a small influence on the surface CO_2 ocean response in the models (i.e. leading to slightly higher $fCO_{2\text{-ocean}}$ trend in the models compared to the trend in the real-world). If to a first degree, the atmospheric xCO_2 is approximated to $fCO_{2\text{-atmosphere}}$ and if the CO_2 system is assumed to be linear, an overestimation of 0.03 µatm·yr^{-1} in the $fCO_{2\text{-atmosphere}}$ in the models would correspond to an overestimation of 0.03 µatm·yr^{-1} in $fCO_{2\text{-ocean}}$, which corresponds to less than 2% of the model-mean $fCO_{2\text{-ocean}}$ trend (i.e. 1.90 µatm·yr^{-1}). As such, the significant bias on the $fCO_{2\text{-ocean}}$ trend in the CMIP5 models is not due to minimal and indistinguishable differences between the atmospheric xCO_2 products.

5.2.2 Model-observation discrepancy explained by observational uncertainty?

Here, the sensitivity of the number of CMIP5 models used to perform the CMIP5-based uncertainty assessment is explored. Indeed, the trend uncertainty assessment, which was used to constrain the observation-based trend estimate, relied on the CMIP5-based MLR analyses and specifically, on the comparison between the trend using the basin-wide $fCO_{2\text{-ocean}}$ that was predicted by the MLR (i.e. $\Gamma_{\text{MLR-predicted}}$) and the trend using the basin-wide $fCO_{2\text{-ocean}}$ that was provided by the true model field (i.e. $\Gamma_{\text{model-true}}$). The trend uncertainty was calculated by the standard deviation of the difference between $\Gamma_{\text{MLR-predicted}}$ and $\Gamma_{\text{model-true}}$ across the 19 CMIP5 models (Equations 4.7, 4.8; see Section 4.3.2 for more details on the trend assessment). As such, the sensitivity analysis on the trend uncertainty will allow us to test whether the significant difference between the $fCO_{2\text{-ocean}}$ observation-based trend and the model-truth CMIP5 trends remains robust if we were to determine the observation-based $fCO_{2\text{-ocean}}$ trend's uncertainty using fewer models in the MLR analysis and different combinations of models.
The sensitivity analysis calculates the trend uncertainty (based on Equation 4.7) using a total number of models *k* that varies from 2 to 18 and considering the different possible combinations C_k^M of those models:

$$C_k^M = \frac{M!}{k!(M-k)!} \tag{5.1}$$

where *M* is the total number of models available (i.e. 19) and *k* the number of selected models that are constituting a new ensemble and used to calculate the uncertainty (i.e. from 2 to 18). The standard deviation (i.e. trend uncertainty) calculations are repeated for each incremented ensemble size and for each possible combination of models.
The observation-based $fCO_{2\text{-ocean}}$ trend with the uncertainties resulting from the sensitivity analysis (i.e. $1.47 \pm (1, 2, 3)\ \sigma_{C_k^M}$ µatm·yr^{-1}) are compared to the CMIP5 model ensemble of $fCO_{2\text{-ocean}}$ trends with an unpaired two-samples left-tailed t-test, for each of the possible combinations (Figure 5.3). The t-test statistics return that no matter how many and which models were selected to compute the trend uncertainty, the $fCO_{2\text{-ocean}}$ observation-based trend is always significantly smaller than the CMIP5 trends, at the 5% significance level (Figure 5.3). The sensitivity analysis therefore indicates that the CMIP5 models robustly

overestimate the observation-based $fCO_{2\text{-ocean}}$ trend.

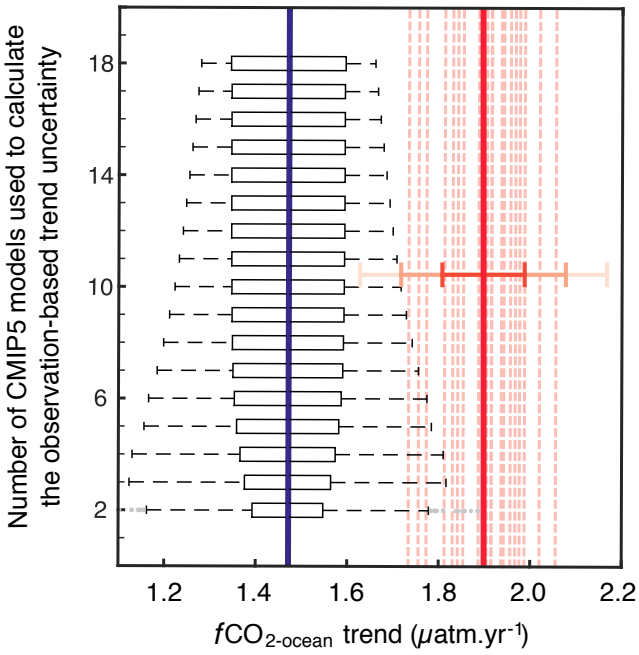

Figure 5.3: A sensitivity analysis on the fCO_2 trend uncertainty. $fCO_{2\text{-ocean}}$ trends for the 19 CMIP5 models (orange) and from the observation-based MLR results (blue). Each box plot contains the $\pm$ (1, 2, 3) $\sigma_{C_k^M}$ trend uncertainties, relative to the observation-based trend, using all the possible combination of results from the CMIP5-based MLR analysis. The y-axis indicates the total number of models that were included in the trend uncertainty analysis (Equations 4.7 and 4.8). Trends are calculated for the North Atlantic and for the period 1992-2014.

5.2.3 A model bias due to internal variability?

One of the major challenge when interpreting the time-varying behaviour in a model ensemble against observations on decadal to multi-decadal timescales is that the model ensemble and the real-world could be experiencing different phases of internal variability, and as such, one may ask: Is the model ensemble substantially different from the observations because it does not capture the component of natural variability sampled by the real-world? To answer this question, the internal variability in the model ensemble and in the real-world should be quantified. However, the interpolated $fCO_{2\text{-ocean}}$ observational record in which we have confidence (i.e. the 23-year long 1992-2014 period), is of insufficient length to capture the multi-decadal modes of variability within the CO_2 system (McKinley et al., 2011). As such, the internal variability is studied in the CMIP5 models through two analyses: the first one

explores the fact models might be in a slightly different phase of variability compared to the real-world's variability in 1992-2014 by calculating additional $fCO_{2\text{-ocean}}$ trends in the models over intervals of 23 years around the period of study, and the second one quantifies the internal variability using the preindustrial experiment runs.

Are model trends sensitive to the period of study?

To test whether or not the CMIP5 $fCO_{2\text{-ocean}}$ trends result from a high frequency signal of internal variability, the models' $fCO_{2\text{-ocean}}$ trends on intervals of 23 years of either side of the period of study (i.e. 1992-2014) are analysed. The North Atlantic $fCO_{2\text{-ocean}}$ trends are calculated over four regular periods of 23 years (1990-2012, 1991-2013, 1993-2015 and 1994-2016) for each of the 19 CMIP5 models (represented by the grey histogram in Figure 5.4). Although the analysis could be extended to additional intervals of 23 years (e.g. 1989-2011, 1995-2017), the atmospheric CO_2 concentrations that those intervals are characterised by are likely to be substantially different from the atmospheric CO_2 concentrations that surface waters of the period 1992-2014 are experiencing, and the trends calculated over those intervals would wrongly suggest that the CMIP5 models in 1992-2014 are capturing a phase of high internal variability (by falling outside of the of 1992-2014 trend ensemble), while in fact they are experiencing different forced signals. While the chosen four intervals (i.e. 1990-2012, 1991-2013, 1993-2015 and 1994-2016) for the present analysis does not allow us to conduct a robust study on the sensitivity of the inter-annual variability in the models, the 76 additionally generated $fCO_{2\text{-ocean}}$ trends across the 19 CMIP5 models provides a first indication whether the year to year variability may be responsible for the discrepancy between the real-world and the models.

We find that the modelled $fCO_{2\text{-ocean}}$ trends calculated for the four intervals (i.e. 1990-2012, 1991-2013, 1993-2015 and 1994-2016) present a similar amplitude as the modelled trends generated for the period of study 1992-2014 (the histogram is within the model-mean and dispersion range in Figure 5.4). As such, the inter-annual variability captured within intervals of 23 years likely does not explain the CMIP5 models' overestimation in the $fCO_{2\text{-ocean}}$ trends.

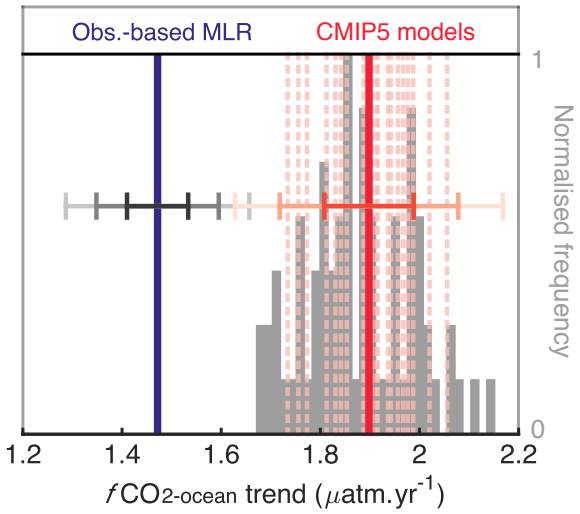

Figure 5.4: The role of inter-annual variability in the CMIP5 models $fCO_{2-ocean}$ trends. The histogram gathers all the $fCO_{2\text{-ocean}}$ trends calculated over the periods 1990-2012, 1991-2013, 1993-2015 and 1994-2016 for the 19 CMIP5 models, to which the frequency has been normalised from 0 to 1. Annually area-weighted averaged North Atlantic $fCO_{2\text{-ocean}}$ over the period 1992-2014 and resulting linear trends. For further figure descriptions, c.f. Figure 5.1.

While the approach used here gives a first indication that the year to year internal variability does not seem to notably impact the CMIP5 model trends evaluation, a more thorough internal variability analysis should be carried out, particularly using preindustrial control runs, which specifically do not experience the increase in atmospheric CO_2 concentrations.

Is the model evaluation sensitive to the size of the model ensemble?

The study here tackles the question: Is the ensemble of 19 CMIP5 models large enough to capture the models' internal variability and therefore provide a model evaluation that reflect the overall forced signal rather than the unforced signal? For instance, if the model ensemble evaluation was based on the model-mean and standard deviation calculated from only 3 CMIP5 models instead of 19, the results might substantially vary depending on which CMIP5 models were selected. In an illustrative example, the mean and standard deviation across 3 $fCO_{2\text{-ocean}}$ model trends could be 1.94 ± 0.18 µatm·yr^{-1} (Figure 5.5a) or 1.82 ± 0.06 µatm·yr^{-1} (Figure 5.5b). While the first example (i.e. 1.94 ± 0.18 µatm·yr^{-1}) would suggest that the models present little agreement among themselves and could potentially present a non-significant difference with the observation-based estimate, the second example (i.e. 1.82 ± 0.06 µatm·yr^{-1}) would suggest that the models consistently overestimate the observation-based value. The second example, which is characterised by a

small standard deviation (i.e. 0.06 µatm·yr^{-1}), might under-sample the models' internal variability and the resulting model evaluation might therefore not be appropriate. As such, the illustrative example (Figure 5.5) suggests that an ensemble of 3 CMIP5 models might not be large enough to provide a consistent model evaluation, and highlights the impact of the models' internal variability when evaluating model trends. However, without a precise quantification of the internal variability existing in the $fCO_{2\text{-ocean}}$ over 23-year long intervals, it is not possible to precisely assess whether or not the 3 CMIP5 models ensemble under-sample the models' internal variability. The present study therefore aims to (1) quantify the range of the models' internal variability in 23-year long intervals using the preindustrial $pCO_{2\text{-ocean}}$ fields, and (2) quantify the minimal size of the CMIP5 model ensemble necessary to consistently capture models' internal variability and therefore provide consistent model evaluation analyses. For clarity purposes, the standard deviation across the $fCO_{2\text{-ocean}}$ trends calculated from the forced model simulations over the period 1992-2014 (i.e. Table 5.1) is hereinafter referred as σ_{forced}, and the standard deviations across the $pCO_{2\text{-ocean}}$ trends calculated from the control runs over the possible 23-year intervals are hereinafter referred as $\sigma_{unforced}$. In the context of this study, the difference between $pCO_{2\text{-ocean}}$ and $fCO_{2\text{-ocean}}$ is assumed to be negligible (Appendix A.1).

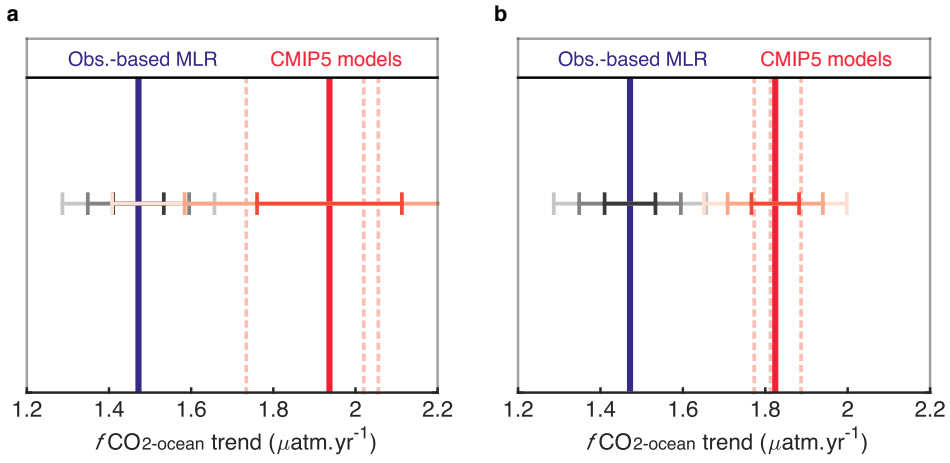

Figure 5.5: Illustrative example of a model evaluation based on 3 CMIP5 models. Model evaluation based on 3 different CMIP5 models where the model-mean and standard deviation in **a**, is 1.94 ± 0.18 µatm·yr^{-1} and in **b**, is 1.82 ± 0.06 µatm·yr^{-1}. For further figure descriptions, c.f. Figure 5.1.

The North Atlantic internal variability (i.e. the unforced variability) in the models' surface $pCO_{2\text{-ocean}}$ contained in 23-year long intervals is quantified by using the models'

preindustrial experimental runs, which describe the climate system without anthropogenic forcing (e.g. the atmospheric CO_2 concentrations are set to preindustrial values). Out of the 19 CMIP5 models used in the model trend evaluation study, 14 provided the preindustrial $pCO_{2\text{-ocean}}$ (Table A.3). The output from these models was subjected to data processing as described in Appendix A.1, and the North Atlantic extracted, and shelf waters (shallower than 1,000 m) removed (Amante and Eakins, 2015). Since the 14 CMIP5 models' control simulations were run over different time lengths (from 240 to 1,000 year-long periods), the first 240 years in each of the control simulations were extracted for consistency purposes. Over 240 years, 217 possible 23-year long continuous intervals are defined, leading to the generation of 217 $pCO_{2\text{-ocean}}$ linear trends for each of the 14 CMIP5 control simulations (histogram on Figure 5.6a).

The standard deviation $\sigma_{unforced}$ of the $pCO_{2\text{-ocean}}$ trends calculated across all the possible 23-year long intervals and across the available control simulations (a total of 217×14 = 3,038 trends) equals 0.036 µatm·yr^{-1} (corresponding to the dark green shading in Figure 5.6a), which therefore corresponds to the amplitude of the unforced variability in the models. The internal variability in the surface $fCO_{2\text{-ocean}}$ trend over a 23-year long interval (up to 3 $\sigma_{unforced}$ = 0.108 µatm·yr^{-1}), as simulated by the CMIP5 models, is about four times smaller than the difference between the CMIP5 model-mean and the observation-based $fCO_{2\text{-ocean}}$ trend (i.e. 1.90 - 1.47 = 0.43 µatm·yr^{-1}), suggesting that the systematic trend overestimation in the CMIP5 models cannot be explained by the models' internal variability. To further support this statement, additional analysis is undertaken to justify that the size of the model ensemble is wide enough to capture most of the internal variability, as simulated by the CMIP5 models.

If the standard deviation σ_{forced} across an ensemble of forced $fCO_{2\text{-ocean}}$ modelled trends is larger than the $2\sigma_{unforced}$ of internal variability's range (i.e. 2×0.036 = 0.07 µatm·yr^{-1}), it suggests that the ensemble is wide enough to captures 95% of internal variability's range (for normally distributed data) and that the evaluation of the CMIP5 models constituting that ensemble most likely describes an evaluation of a mean state that is beyond internal variability. Using the above illustrative example, the standard deviation σ_{forced} calculated across an ensemble of 3 CMIP5 models was 0.06 µatm·yr^{-1} (Figure 5.5b), which by being smaller than the 95% range of internal variability of 0.07 µatm·yr^{-1} (i.e. $2\sigma_{unforced}$),

suggests that the 3 selected CMIP5 models most likely describe a phase of internal variability rather than the forced signal, leading to inconclusive model evaluation. Nevertheless, due to the increase in atmospheric CO_2 concentrations since the beginning of the industrial era, the surface $fCO_{2\text{-ocean}}$ is at present day higher than during preindustrial times, leading to the surface ocean being less buffered and potentially more variable than prior the existence of anthropogenic forcing. As such, the dispersion of the trends simulated by the preindustrial runs are not necessarily directly comparable with the dispersion of the trends simulated in an anthropogenically-forced climate.

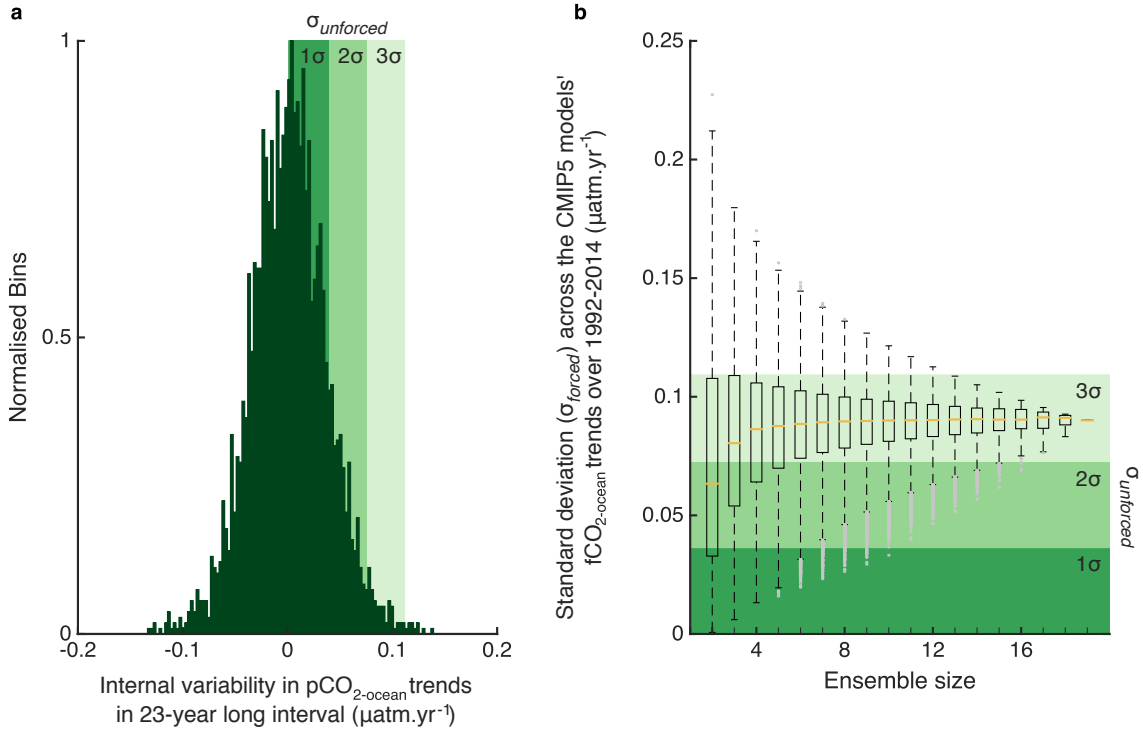

Figure 5.6: Quantifying the North Atlantic internal variability in modelled trend. **a**, Histogram of the $pCO_{2\text{-ocean}}$ linear trends calculated over 23-year running intervals, using the first 240 years of the preindustrial experiment outputs, available for 14 CMIP5 models (Table A.3). The dark to light green bands correspond to 1, 2 and 3$\sigma_{unforced}$ of those $pCO_{2\text{-ocean}}$ trends, respectively, and which are reported into **b**. **b**, Box plot of the standard deviations σ_{forced} across the $fCO_{2\text{-ocean}}$ linear trends calculated in 1992-2014 (using the historical and RCP8.5 experiments) when considering different ensemble sizes (i.e. the total number of models included in the standard deviation calculations) and all the possible combinations of models given an ensemble size.

Now that the amplitude of internal variability in trends over 23-year long intervals has been quantified, the minimum number of models that are required to create an ensemble that captures 95% of internal variability (i.e. 2σ = 0.07 µatm·yr^{-1} for normally distributed data)

can be determined. For each possible ensemble, whose size varies from 2 to 18 CMIP5 models (19 being the ensemble used in Figure 5.1) considering all possible model combinations (Equation 5.1), the new standard deviation σ_{forced} is calculated following:

$$\sigma_{forced} = \sqrt{\sum_{m=1}^{k} \frac{\left(\Gamma_m - \frac{1}{k}\sum_{m=1}^{k}\Gamma_m\right)^2}{k-1}} \tag{5.2}$$

where Γ_m corresponds to the $fCO_{2\text{-ocean}}$ trends calculated over the period 1992-2014 for the CMIP5 model *m* selected within the combination of models forming the new ensemble of size *k*, following Equation 5.1. For example, for an ensemble of three CMIP5 models (i.e. *k*=3), there are 969 possible combinations of 3 models among 19 models (Equation 5.1), leading to 969 values of σ_{forced} (reported as a box plot on Figure 5.6b).

When the ensemble contains less than 17 CMIP5 models, the amplitude of σ_{forced} (i.e. y-axis in Figure 5.6b) can be of a same amplitude as $2\sigma_{unforced}$, depending on the model configurations. For example, for an ensemble size of 16 CMIP5 models, few outlier σ_{forced} values (represented by the grey dots in Figure 5.6b) are within the $2\sigma_{forced}$ shading. As such, the evaluation of an ensemble of less than 17 CMIP5 models might, depending on the selected model combination, not adequately capture interval variability (e.g. McKinley et al., 2016). The smaller the ensemble size, the more likely it can under-sample the unforced variability. When the ensemble contains all the available CMIP5 models (i.e. *k*=19), the amplitude of σ_{forced}, which equals 0.09 µatm·yr^{-1}, is larger than $2\sigma_{unforced}$ (i.e. 0.07 µatm·yr^{-1}), which suggests that the ensemble of 19 forced model runs (as used in this study) is large enough to sample the models' unforced variability captured within an interval of 23 years.

While the $fCO_{2\text{-ocean}}$ trends in the 19 ensemble members seems to predominantly capture the forced change rather that the models' unforced variability, it cannot be ruled out that the observation-based $fCO_{2\text{-ocean}}$ trend estimate may describe a phase of decadal internal variation, that is not simulated by the models. For instance, the North Atlantic CO_2 sink might be experiencing the influence of the hiatus on the global mean surface temperature, which occurred in 1998-2013 (e.g. Xie and Kosaka, 2017) and was not simulated by any of the CMIP5 models (Marotzke and Forster, 2015). While the analysis carried out in Chapter

4 on the drivers of the observation-based $fCO_{2\text{-ocean}}$ trend (c.f. Figure 4.5b) identified SST as the least contributing variable (relative to the other explanatory variables), the mechanisms behind the hypothetical influence of the global warming hiatus on the North Atlantic biogeochemistry could describe the increase in the CO_2 uptake in that basin. Indeed, a decrease in the surface warming (as described by the observed global hiatus) could lead to (1) an increase in the solubility of CO_2 into seawater (solubility being negatively correlated to temperature), and (2) a decrease in stratification which would enhance vertical ocean mixing and as such, potentially stimulate primary production through the input of nutrient-enriched waters to the surface; both effects stimulating the air-sea CO_2 flux (Chapter 1). Nevertheless, a preliminary analysis (not displayed) suggested that some of the CMIP5 models captured the observed amplitude of the North Atlantic SST gradient during the hiatus period (i.e. 1998-2013), which further suggest that a biogeochemical effect inherent to the CMIP5 models might explain the discrepancy between the models and the real-world $fCO_{2\text{-ocean}}$ trend (further discussion and analysis on the reasons of the systematic model bias in Section 5.3).

As discussed in Chapter 1, the observational-based study of McKinley et al. (2011) identified that in the North Atlantic, the long-term trend takes 25 years to emerge from the variability occurring on decadal timescales. If the 25 years threshold identified by McKinley et al. (2011) is proven to be a solid dividing line between the decadal internal variability and the emergent anthropogenic forcing, the observation-based $fCO_{2\text{-ocean}}$ trend estimate calculated over the 23-year long 1992-2014 period should then capture decadal internal variability rather than the forced change. However, during the observation-based MLR analysis conducted in Chapter 4 (c.f. Section 4.4), the atmospheric xCO_2 was identified as the dominating contributor to the surface $fCO_{2\text{-ocean}}$ trend, suggesting that the forced change dominates the $fCO_{2\text{-ocean}}$ signal over the period 1992-2014. Nevertheless, without a well-distributed observational CO_2 record that is long enough to capture the decadal and multi-decadal modes of variability, the cause(s) behind trends in the real-world CO_2 system cannot be fully investigated (McKinley et al., 2011).

In summary, the discrepancy between CMIP5 models and observation-based $fCO_{2\text{-ocean}}$ trends (Figure 5.1) most likely does not arise from the subtle differences in the atmospheric

$x\mathrm{CO}_2$ forcing, the trend uncertainty assessment used to constrain the observation-based estimate and internal variability in the model ensemble (the ensemble being wide enough to capture the models' internal variability); but likely arises from the failure of the models to reproduce the real-world's internal variability and/or the real-world's response to anthropogenically forced increase in atmospheric CO_2 concentrations.

5.3 Investigating the mechanisms explaining the bias in the CMIP5 models

While the CMIP5 models are the most advanced tools available at this time to explore the Earth's climate response to anthropogenic forcing, and the coordination of their simulations to provide multi-model ensembles is invaluable in avoiding drawing erroneous model-specific conclusions, such multi-model analysis lead to challenges. Different modelling groups will have made different assumptions in building and setting up the model, for example with different initial conditions and spinup procedures (Séférian et al., 2016). As such, to understand and identify the potential mechanisms responsible for the systematic bias in the $f\mathrm{CO}_{2\text{-ocean}}$ trends in the CMIP5 models, five ocean-only simulations, using a single model, are performed. The five simulations are identically initialised with observation-based products, but are forced with five different surface condition products spanning the interval of interest, which allow us to investigate the impacts of the initialisation and of the surface forcing fields on the simulated $f\mathrm{CO}_{2\text{-ocean}}$ trends.

After having presented the simulations set-up (Section 5.3.1) and addressed the model drift within each of the five simulations (Section 5.3.2), the importance of the initialisation of biogeochemical fields in simulating the North Atlantic suface $f\mathrm{CO}_{2\text{-ocean}}$ trends is discussed (Section 5.3.3).

5.3.1 Simulation set-up

Simulations were performed using a 1° global ocean-only physical-biogeochemical model, the 75 vertical levels version of the GO5.0 (Megann et al., 2014) configuration of the Nucleus for European Modelling of the Ocean (NEMO) hydrodynamic model (Madec, 2008), cou-

pled with v4.1 of the Los Alamos National Laboratory sea ice model (CICE) model (Hunke et al., 2010), and the Hadley Centre Ocean Carbon Cycle (HadOCC) (Palmer and Totterdell, 2001) biogeochemical model. The HadOCC ocean carbon cycle model follows a NPZD configuration, which is defined by six state variables (including DIC and TA) and by fixed carbon:nitrogen ratios. HadOCC includes biological processes such as primary production, respiration, grazing, detritus sinking and remineralisation at depth (Palmer and Totterdell, 2001).

All NEMO fields were initialised to zero, except for temperature and salinity which were taken from the EN4 objective analysis v4.1.1 (Good et al., 2013; Gouretski and Reseghetti, 2010). For HadOCC, initial conditions for nutrients were taken from the World Ocean Atlas (WOA) climatology (Garcia et al., 2010), for DIC and TA from the GLODAP climatology[2] (Key et al., 2004), and for phytoplankton, zooplankton and detritus from the end of a previous simulation (Ford et al., 2012). The atmospheric pCO_2 values were prescribed using globally and monthly averaged surface data based on observations (Dlugokencky and Tans, 2016).

The five simulations were run from 3^{rd} January 1979 to 31^{st} December 2014, and were forced by prescribing the surface conditions from five different sources (one source for each experiment). The prescribed surface forcings include air temperature, snowfall, specific humidity, vector winds, precipitation, downwelling shortwave and longwave radiation. The motivation behind running five simulations forced with five different surface conditions was to test whether the $fCO_{2\text{-ocean}}$ simulated trend was highly impacted by the atmospheric forcing (heat, moisture and momentum fluxes) or by the initialisation of the ocean variables which remained similar across the simulations. The first simulation used daily surface conditions from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis (Dee et al., 2011) and is hereinafter referred as the "ERA-Interim forced simulation"[3]. The remaining four simulations used daily surface conditions from four CMIP5 model outputs, specifically from the GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and CanESM2 (whose data processing is described in Appendix A.1), and are hereinafter referred

[2]GLODAPv1 was used because this version is more representative of the 1980s, period over which the simulation initialisation is made, than the second updated version.

[3]A separate simulation was computed using 3-hourly fluxes also from ECMWF ERA-Interim reanalysis, but this change in forcing frequency did not significantly impact the $fCO_{2\text{-ocean}}$ trend. For consistency with the CMIP5-forced simulations, which are forced using daily modelled fields, the daily ERA-Interim forced simulation is kept for the present study.

as the "CMIP5-forced simulations". The four CMIP5 models were chosen to approximately span the $fCO_{2\text{-ocean}}$ CMIP5 model behaviours (Table 5.1).

For each of the five simulations, the surface $fCO_{2\text{-ocean}}$ outputs were (1) regridded into a $1^\circ \times 1^\circ$ regular grid using the bilinear interpolation within the Climate Data Operator (CDO) package (http://www.mpimet.mpg.de/cdo), (2) extracted for the open waters of the North Atlantic (the shelf waters above 1,000 m depth were removed using ETOPO1, Amante and Eakins, 2015), (3) monthly averaged (as the outputs were save at a daily frequency) and (4) averaged into annual area-weighted means; leading for each simulation to a $fCO_{2\text{-ocean}}$ annual time series from 1979 to 2014. For each of the five annual $fCO_{2\text{-ocean}}$ annual time series, the linear trend was calculated over the period of interest 1992-2014. As such, the first 13 years of the simulations (i.e. from 1979 to 1991) correspond to the spinup period. While a 13-year spinup is inevitably of insufficient length for the model to reach equilibrium, an analysis of the model fields suggested that the large initial drifts had settled down during this period (the initial perturbation settles after $\sim$3-4 years, example in Figure 5.7), certainly for the surface processes which we are most interested in over the timescales of this study, and which are most driven by the atmospheric forcing. As such, we are satisfied that the hindcast is sufficiently able to reproduce the mean state and observed variability of the ocean (Ford and Barciela, 2017). Furthermore, we require a short spinup to avoid the DIC and TA fields drifting too far from their initialised observed state, and therefore to test the carbon cycle behaviour when the model's DIC and TA are close to the observed (initialised) state on the simulated surface $fCO_{2\text{-ocean}}$ trends.

From the five simulated $fCO_{2\text{-ocean}}$ trends, the following questions are addressed: Does the $fCO_{2\text{-ocean}}$ trend obtained from the ERA-Interim forced simulation match the observation-based trend estimate better than do any of the CMIP5 models? Do the $fCO_{2\text{-ocean}}$ trends obtained from CMIP5-based forced simulations provide an improvement compared to their respective CMIP5 trends and as such, is the observation-based initialisation the key feature which allows us to simulate a bias-free $fCO_{2\text{-ocean}}$ trend? To suitably address these questions, the impact of the potential model drift on the simulated $fCO_{2\text{-ocean}}$ trends must be quantified.

5.3.2 Model drift

To identify whether the simulated $fCO_{2\text{-ocean}}$ trends are a result of potential model drift and thus to provide a meaningful comparison between the simulated, the observation-based and the CMIP5 models $fCO_{2\text{-ocean}}$ trends, additional simulations for which the atmospheric CO_2 concentration was held constant were run. In constant atmospheric CO_2 concentration conditions, if the surface $fCO_{2\text{-ocean}}$ remains approximately constant, the model is considered as non-drifting. Here, the method (1) quantifies the model-drift in each of the five simulations, and (2) accordingly removes the model drift from the simulated $fCO_{2\text{-ocean}}$ trends.

Quantification of the model drift

To quantify the model drift in each of the five ocean-only simulations, fifteen additional simulations with prescribed constant atmospheric CO_2 mixing ratio, were run; hereinafter referred as the "model-drift simulations". The model-drift simulations all experienced the same constant atmospheric CO_2 mixing ratio of 336.85 ppm, which is the mixing ratio value for the year 1979 (i.e. the starting year of the changing CO_2 runs; Section 5.3.1) (Dlugokencky and Tans, 2016). The fifteen model-drift simulations were divided into five groups, each of them was respectively forced with the daily surface conditions from ERA-Interim, GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and CanESM2. Each group is therefore made of three model-drift simulations, each of which were initialised at three different dates and hence run over three different periods of 36 years: from February 1979[4] to January 2015, from January 1984 to January 2020, and from January 1988 to January 2024. By shifting the atmospheric forcing years (i.e. from 1979 to 1984 and from 1979 to 1988), the set of model-drift simulations per surface conditions products allows us to separate the model drift that is purely due to the model adjusting to the initial conditions from the trend that results from the year to year variability in the atmospheric forcing conditions. Since the ERA-Interim reanalysis was available until 2015, and since two

[4]The outputs of five main simulations (i.e. the ones with increasing atmospheric CO_2) were saved at a daily frequency, which was a particularly time-expensive and not particularly necessary task for the purpose of this study, as we were interested in monthly and annually outcomes. To accelerate the running simulation time, the code was modified to save monthly outputs. Since the ECMWF ERA-Interim field is available from the 2nd January 1979, the monthly-averaged model-drift simulation could not start in January 1979 as it would have required information from December 1978 to start the simulation (unlike the daily-saved simulation which could start on the 3rd of January 1979 and use the information from the previous time-step (i.e. the 2nd of January 1979) to initiate the simulation).

ERA-Interim-forced model-drift simulations required data until 2020 and 2024, the ERA-Interim surface conditions from the period January 1983 to January 1988 were used to cover the period[5] from January 2015 to January 2020, and the data from the period January 1983 to January 1992 were used to cover the period from January 2015 to January 2020. Since the CMIP5 models' surface conditions are known in the future, the daily models' surface conditions from the scenario RCP8.5 were used up to 2024 for the CMIP5-forced model drift simulations.

For each of the fifteen model-drift simulations, the $fCO_{2\text{-ocean}}$ outputs followed the same data processing steps as for the five changing atmospheric CO_2 runs (Section 5.3.1). From the fifteen annual time series of the North Atlantic area-weighted means of surface $fCO_{2\text{-ocean}}$ (Figure 5.7), the linear trends were calculated from the year 14 (i.e. the end of the spinup phase) to the year 36 (i.e the end of the simulation), corresponding to the periods 1992-2014, 1997-2019 and 2001-2023 and for the simulations started in 1979, 1984 and 1988, respectively (Table 5.7). The amplitude of the model drift in the surface $fCO_{2\text{-ocean}}$ varies across the differently forced and initiated simulations, suggesting that the different mixing and solubility have an impact of the resulting $fCO_{2\text{-ocean}}$ (Table 5.7). Overall, the ocean-only model simulations experience significant negative surface $fCO_{2\text{-ocean}}$ trends, except when forced with the HadGEM2-ES surface forced conditions starting in 1984 and 1988, at the 5% significant level (Table 5.7). As such, in the case where the surface $fCO_{2\text{-ocean}}$ is increasing over the period 1992-2014 in the changing atmospheric CO_2 runs, the identified model-drift would actually mean that the surface $fCO_{2\text{-ocean}}$ is actually experiencing a stronger rate of increase compare to the non-corrected model-drift results.

[5]The reason for choosing to repeat the data from 1983 to 1998 to cover the period from 2015 to 2020 relies on leap years constraints. Indeed, since 2016 is a leap year, the interval was chosen so that 1984 which is the also a leap year matched that requirement. The leap year 1980 was not selected because the previous year 1979 was not completely available (the ERA-Interim data for 1979 were available from the 3rd of January).

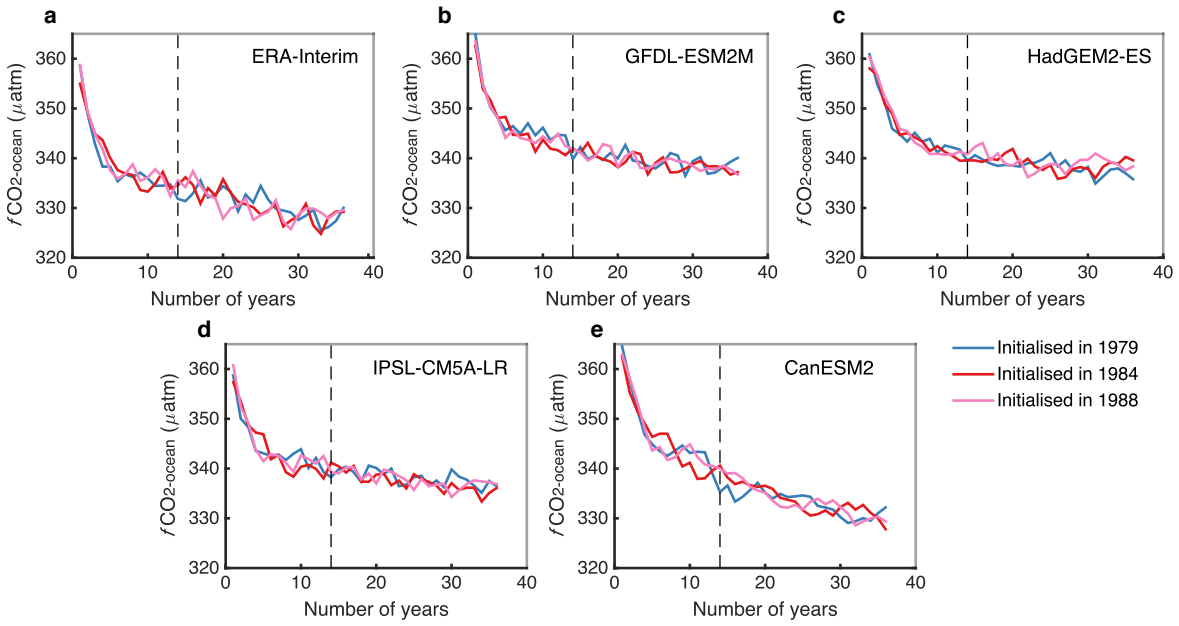

Figure 5.7: Quantification of the model-drift in the North Atlantic surface $fCO_{2-ocean}$. North Atlantic (area-weighted) $fCO_{2\text{-ocean}}$ annual time series calculated from the simulations for which the atmospheric CO_2 mixing ratio was held to 336.85 ppm and were forced with **a**, ERA-Interim, **b**, GFDL-ESM2M,**c**, HadGEM2-ES, **d**, IPSL-CM5A-LR and **e**, CanESM2 surface conditions. For each of the five surface conditions products, three simulations were made with three different starting and therefore initialisation date: January 1979 (blue), January 1984 (red) and January 1988 (pink). The first 13 years (left to the dashed line) correspond to the spinup phase, which are excluded from the model-drift assessment. The model-drift is quantified using the data on the right of the dashed line. For each simulation, the corresponding model-drift (i.e. linear trend) is indicated in Table 5.2.

forcing field / sim. start year	**ERA-Interim**	**GFDL-ESM2M**	**HadGEM2-ES**	**IPSL-CM5A-LR**	**CanESM2**
1979	**-0.29 ± 0.06**	**-0.11 ± 0.04**	**-0.15 ± 0.03**	**-0.14 ± 0.04**	**-0.29 ± 0.04**
1984	**-0.37 ± 0.06**	**-0.18 ± 0.04**	-0.08 ± 0.05	**-0.24 ± 0.03**	**-0.42 ± 0.03**
1988	**-0.29 ± 0.07**	**-0.19 ± 0.04**	-0.10 ± 0.05	**-0.15 ± 0.04**	**-0.44 ± 0.04**

Table 5.2: Trends in the North Atlantic surface $fCO_{2-ocean}$ model drift. Linear trends ($\mu atm \cdot yr^{-1}$) of the model drift in the North Atlantic $fCO_{2\text{-ocean}}$ calculated from year number 14 to 36 (Figure 5.7) for the fifteen model drift simulations, a period that corresponds to a 23-year period long similar to the interval of study 1992-2014 (for the changing atmospheric CO_2 runs). The uncertainties correspond to the standard error returned by each linear fit. Significant trends are in bold, at the 5% significance level.

Correction of the model-drift in the simulated $fCO_{2-ocean}$ trends

To correct the model drift in each of the five North Atlantic $fCO_{2\text{-ocean}}$ trends simulated by the atmospheric varying xCO_2 runs, a series of three main steps were carried out. Here is an example for the ERA-Interim forced simulations:

1. For the model-drift simulation started in 1979, a linear regression was fitted between the years 14 and 36 of the simulation (thick blue line in Figure 5.8a). The linear trend and standard error σ values returned by the linear fit (-0.29 $\pm$ 0.06 µatm·yr^{-1}; Table 5.2) were used to construct six additional linear time series (dashed blue lines in Figure 5.8a). The first time series was constructed so that it had a trend equalling the + 1 σ limit from the main fit: -0.29 + 0.06 µatm·yr^{-1} = -0.23 µatm·yr^{-1}. The second time series was similarly constructed but from the - 1 σ limit from the main fit, which therefore had a trend of -0.35 µatm·yr^{-1}. The third to the sixth time series were finally constructed from the $\pm$ 2, 3 σ limits from the main fit, leading to four time series with the trends of -0.47, -0.41, -0.17 and -0.11 µatm·yr^{-1}. As such, there are a total of seven fitted time series for the model-drift simulation starting in 1979. These steps were repeated for the model-drift simulation starting in 1984 and in 1988 (Figure 5.8b and c, respectively), leading to a total of 21 model drift estimates for the ERA-Interim forced model drift simulations (all the lines in Figure 5.8a-c)

2. The 21 model drift estimates were subtracted one at a time from the $fCO_{2\text{-ocean}}$ time series over the period 1992-2014 for the changing atmospheric CO_2 run (i.e. dashed line in Figure 5.8d, whose linear trend is 1.29 $\pm$ 0.06 µatm·yr^{-1}, where 0.06 µatm·yr^{-1} corresponds to the linear fit's standard error), leading to 21 possible corrected $fCO_{2\text{-ocean}}$ time series (i.e. coloured lines in Figure 5.8d). Each time series is presented as an anomaly plot relative to its corresponding $fCO_{2\text{-ocean}}$ value in 1991.

3. A linear fit was applied to each of the 21 corrected $fCO_{2\text{-ocean}}$ time series (i.e. coloured lines in Figure 5.8e), whose mean and standard deviation define to the model-drift corrected final North Atlantic $fCO_{2\text{-ocean}}$ trend value simulated by the ERA-Interim forced simulation (i.e. solid black line and error bars Figure 5.8e): 1.61 $\pm$ 0.13 µatm·yr^{-1} (Table 5.3).

The above steps were repeated across each of the CMIP5-forced simulations and whose North Atlantic surface $fCO_{2\text{-ocean}}$ trends prior and after the model drift correction are provided in Table 5.3.

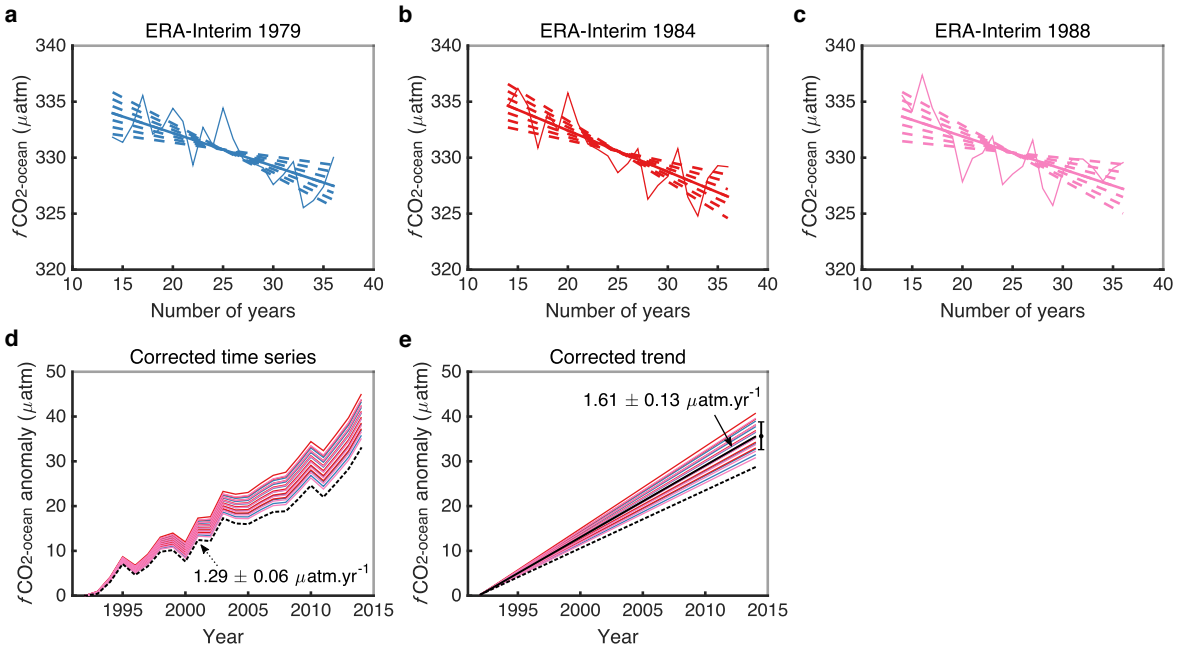

Figure 5.8: Example of the model drift correction steps for the ERA-Interim forced simulation. The figure description is detailed in the steps 1 to 3 above. The coloured lines in **d** and **e** refer to the different starting year of the model drift simulations, as indicated in **a**, **b** and **c**. All annual values were computed for the area-weighted North Atlantic, without shelf waters.

Forcing field	**Non drift corrected trends** (μ**atm**·**yr**$^{-1}$)	**Drift-corrected trends** (μ**atm**·**yr**$^{-1}$)
ERA-Interim	1.29 ± 0.06	1.61 ± 0.09
GFDL-ESM2M	1.36 ± 0.05	1.52 ± 0.09
HadGEM2-ES	1.43 ± 0.04	1.54 ± 0.09
IPSL-CM5A-LR	1.25 ± 0.05	1.43 ± 0.04
CanESM2	1.29 ± 0.05	1.68 ± 0.11

Table 5.3: Simulated trends in the North Atlantic surface $fCO_{2-ocean}$. For the the ERA-Interim forced simulation and from the CMIP5-forced simulation, $fCO_{2\text{-ocean}}$ trends calculated over the period 1992-2014 prior and after model drift corrections. For the non drift-corrected trends, the uncertainty corresponds to the standard error returned by the linear fit, while for the drifted-corrected trends, the uncertainty corresponds to the standard deviation across the 21 trends identified from the model-drift analysis (see text).

5.3.3 Simulated trends

Overall, the five simulated North Atlantic $fCO_{2\text{-ocean}}$ trends show good agreement with the observation-based trend and among themselves (potentially less agreement with the CanESM2-forced simulation) (Figure 5.9). By being forced with an observation-based product, the ERA-Interim run provides an improvement in simulating the North Atlantic $fCO_{2\text{-ocean}}$ trend compared to the averaged CMIP5 model (Figure 5.9a). Nevertheless, a realistic forcing field is not a necessary condition for simulating an appropriate North Atlantic $fCO_{2\text{-ocean}}$ trend; the initialisation with observation-based biogeochemical (DIC and TA) and physical (T and S) fields is. Indeed, the CMIP5-forced simulations are performed with non-realistic atmospherical conditions, and yet they provide realistic North Atlantic $fCO_{2\text{-ocean}}$ trend (Figure 5.9b-e). Over the period 1992-2014, the resulting surface $fCO_{2\text{-ocean}}$ trends calculated from the four CMIP5-forced simulations (i.e. using the surface fluxes from GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and CanESM2) are respectively 1.52 ± 0.09 µatm·yr^{-1}, 1.54 ± 0.09 µatm·yr^{-1}, 1.43 ± 0.04 µatm·yr^{-1} and 1.68 ± 0.11 (Table 5.3), corresponding to a respective averaged improvement of 83%, 86%, 92% and 62% relative to the observation-based $fCO_{2\text{-ocean}}$ trend (Figure 5.9b-e). As such, the inability of the CMIP5 models to capture the observation-based trend seems to not arise from changes in the winds, heat or freshwater forcing coming from the atmosphere, but rather from the biogeochemical fields, which the models inadequately generate in response to their biology and circulation (e.g. DIC and TA in Chapter 4) and also potentially from the T and S observation-based initialisation.

While the present study highlights the importance of the initialisation (or correct simulation) of the biogeochemical, T and S fields, we cannot definitely say that this is the cause of the bias in the CMIP5 models because (1) the ocean-only model used here does not include the same amount of complexity as an ESM (e.g. imbalance in the evaporation-minus-precipitation) and as such, the deduced results are not exactly comparable (rather they act as a simplification of the system to help us understand the behaviour in the CMIP5 models), and (2) the observation-based initialisation in the ocean-only model was performed on three groups of variables: temperature/salinity, nutrients and DIC/TA, which could all have played a role in controlling the surface $fCO_{2\text{-ocean}}$ (Chapter 1). However, the substantial systematic TA biases in the CMIP5 models

(at both the surface and at depth) identified in the model evaluation study against GLODAPv2 (Chapter 2, e.g Figure 2.12b) suggest that the initialisation of the TA (and of DIC) is likely to play a key role in improving the simulated $fCO_{2\text{-ocean}}$ trend. While a simulation initialised with the CMIP5 DIC and TA would have allowed us to test this suggestion, the challenges associated with the regridding of the CMIP5 models' DIC and TA fields[6] to the ocean-only model's grid has prevented us to pursue this investigation.

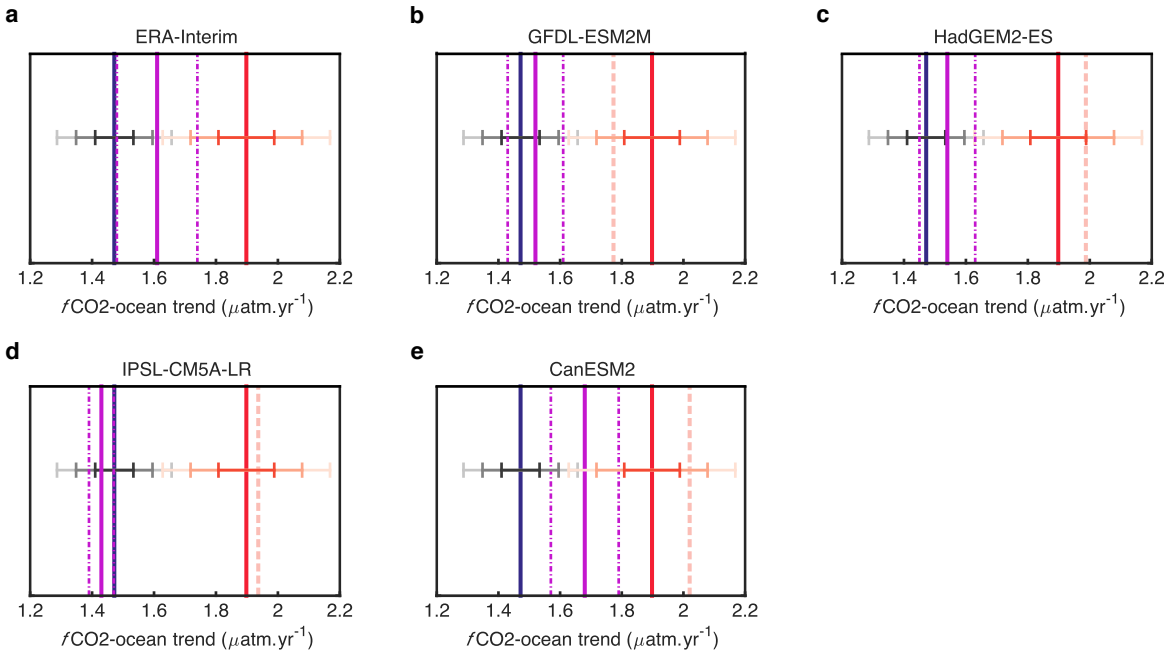

Figure 5.9: The role of biogeochemistry initialisation in reducing model biases on the $fCO_{2-ocean}$ trend. North Atlantic $fCO_{2\text{-ocean}}$ trends (thick pink line) and uncertainty due to inter-annual variability in the atmospheric forcing (dashed pink lines) returned from the ocean-only simulations that were forced with **a**, ERA-Interim, **b**, GFDL-ESM2M, **c**, HadGEM2-ES, **d**, IPSL-CM5A-LR and **e**, CanESM2 daily surface conditions. The orange dashed line in **b**, **c**, **d**, and **e** shows the trend value for the corresponding CMIP5 model (e.g. in **b**, the dashed orange line is the $fCO_{2\text{-ocean}}$ trend calculated from the GFDL-ESM2M model). All simulated trends are here model-drift corrected. For further figure descriptions, c.f. Figure 5.1.

5.4 Summary

While the latest generation of Earth Systems Models (ESMs) are widely used for policy making, and of particular relevance here, to determine the allowable CO_2 emissions to remain below agreed levels global warming, their evaluation typically focuses on how well they repre-

[6] A slight modification of the DIC and TA fields during the interpolation regridding step returned unrealistic $fCO_{2\text{-ocean}}$ values.

sent the climatological state. Since questions asked by policy makers relate to how the system is changing, which are answered by using ESMs, it is important to assess how these models simulate change. By quantifying robust time-varying uncertainties on the surface $fCO_{2\text{-ocean}}$ in the recent past and by investigating the impact of internal variability on the simulated trends, which was found to be minimal, we have been able to identify that the CMIP5 model ensemble underestimated the trend in the North Atlantic CO_2 uptake over the period 1992-2014 by about 3 times relative to the real-world uptake. The potential mechanisms impacting the simulated North Atlantic surface $fCO_{2\text{-ocean}}$ were explored with an ocean-only model, which suggested that the discrepancy between the surface $fCO_{2\text{-ocean}}$ trend in the real-world and in the CMIP5 models arises from inadequacies in their simulation of the background biogeochemical state. Inadequacies in CMIP5 physics and/or biology lead to substantial biases in TA (Chapter 2) that may be the cause for the $fCO_{2\text{-ocean}}$ trend overestimation in the models through its impacts on the buffer capacity (Chapter 1). The present analysis therefore brings into question our ability to predict the decadal-to-centennial evolution of the North Atlantic sink for CO_2 in the future. The model biases identified here and the extension of continuous ocean CO_2 measurements must be addressed if the climate-science community is to provide the best possible guidance on what anthropogenic CO_2 emissions are consistent with agreed atmospheric CO_2 targets.

Chapter 6

Summary & Perspectives

This chapter aims to bring together the main results of this thesis into a summary paragraph (Section 6.1), and explore additional analyses that could help us (1) identify the exact reason(s) for the bias in the CMIP5 models (Section 6.2), (2) understand the change in the CO_2 sink in the world's oceans (Section 6.3), and (3) provide bias-free future predictions for the North Atlantic and global CO_2 oceanic sink (Section 6.4).

6.1 Thesis summary

To study the future behaviour of the carbon reservoirs and determine the threshold of allowable anthropogenic CO_2 emissions given different climate scenarios, models are commonly used. Yet the validation of the models' time-varying CO_2 ocean uptake remains challenging due to very few accurate estimates of the rate of change in the oceanic CO_2 uptake. Indeed, the time-varying CO_2 oceanic uptake is difficult to robustly determine due to the spatially and temporarily incomplete nature of *in-situ* CO_2 measurements. While various gap-filled oceanic biogeochemical products have been developed to provide a global understanding of the marine CO_2 system, gap-filling methods introduce various sources of uncertainty of unknown basin-wide amplitudes (e.g. interpolation error, aliasing effect), which consequently alter our precise understanding of the system and limit the tight constrain of model evaluation. By bringing together the available models from the last IPCC inter-comparison exercise (the CMIP5 multi-model ensemble), with one of the latest released observational CO_2 dataset, the trend and uncertainty in the North Atlantic CO_2 uptake from 1992 to 2014 has been robustly

quantified. The statistical robustness of the trend in the North Atlantic CO_2 uptake has allowed us to (1) provide new fundamental understanding about the operation of this CO_2 sink, (2) identify the mechanisms driving the change, (3) diagnose the rate of change simulated by the CMIP5 ensemble, which was found to unanimously underestimate the observation-based trend by a factor $\sim$3, a discrepancy that is not likely to be due to the models' internal variability but rather by a persistent biased signal, and (4) identify, by building ocean simulations initialised with observational data, that biases in the models ocean biogeochemistry likely impact the response of the North Atlantic CO_2 uptake to rising atmospheric CO_2 concentrations. The results of this thesis therefore illustrate the complexity and multiplicity of factors that influence the ocean sink for atmospheric CO_2, and show what needs to be addressed to improve model projections of ocean carbon uptake.

6.2 Perspectives: further investigation on the reason(s) behind the bias in the CMIP5 models

The simulations carried out in Chapter 5 provided an initial investigation of the reasons behind the model bias, and identified that the initialisation of the concentrations of DIC, TA, nutrients, T and S from observation-based products played a key role in setting the appropriate conditions for the model to simulate the change in the North Atlantic surface CO_2 sink. The North Atlantic DIC and TA model evaluation analysis carried out in Chapter 2 against the observation-based GLODAPv2 product, identified systematic biases in the CMIP5 models' TA, which are likely to be the main reason behind the poor representation of the North Atlantic surface $fCO_{2\text{-ocean}}$ change in the models. To further explore the reasons behind models' discrepancy against observation-based results in the North Atlantic, the models' climatological salinity profiles (Section 6.2.1) and Revelle factor (Section 6.2.2) are studied.

6.2.1 The salinity profiles

The TA biases in the CMIP5 models, identified in the North Atlantic (Figure 6.1a), are likely to arise from the integration, over the spinup, of errors in (1) the air-sea flux of freshwater (diluting or concentrating alkalinity), (2) errors in the models' ocean circulation

field, and (3) errors in the biological vertical movement of alkalinity. The first two points, which are physically-driven mechanisms, will have a similar impact on both alkalinity and salinity. By examining the salinity errors in the CMIP5 models, the physical contribution in the bias in TA can therefore be investigated.

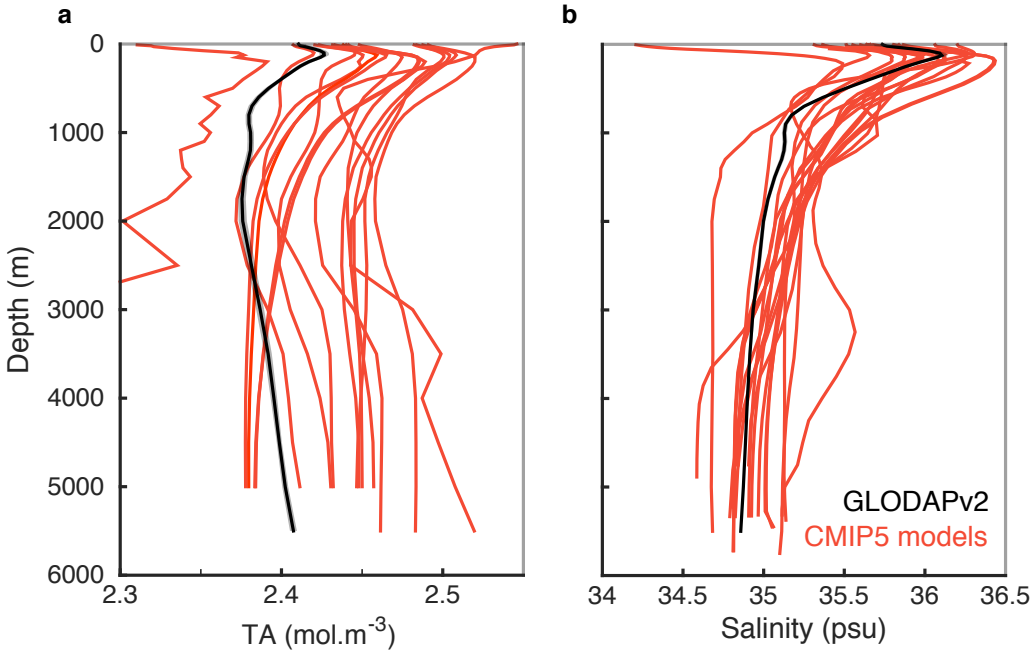

Figure 6.1: Salinity model evaluation at depth. Climatology of the North Atlantic area-weighted **a**, TA (as in Figure 2.10b), **b**, salinity profiles from GLODAPv2 (black) and from 15 and 18 CMIP5 models, respectively (orange; Table A.3, except that the MRI-ESM1 model was not included in **b** due to issues with the time dimension). The interpolation uncertainty provided by GLODAPv2 is plotted as dark shading around the mean profile, but is indistinguishable on those salinity scales. The profiles do not include any shelf waters, which were initially removed using the ETOPO1 bathymetry product (Amante and Eakins, 2015). The models' climatology was computed using the the models' outputs from 1972 to 2013 (as in GLODAPv2).

The North Atlantic (area-weighted) climatological salinity profiles in the CMIP5 models are compared to the observation-based equivalent, provided by GLODAPv2 (Lauvset et al., 2016; Figure 6.1b). Assuming that the GLODAPv2 climatological salinity profile does not contain any major sources of uncertainties (Lauvset et al., 2016), the evaluation of the CMIP5 models (Figure 6.1b) shows that (1) at the near-surface and in the first hundred meters, the CMIP5 models overall capture the amplitude and variations of the salinity signal (at least on a basin averaged basis), and (2) at depth (from about 700 m depth), the models tend to systematically overestimate the observation-based salinity profile. These results suggest that physical processes likely do not explain the models' bias in TA in the upper layer of the

North Atlantic (Figure 6.1a), but potentially explain some of the TA bias further at depth. For example, a weakly simulated freshwater inputs from rivers and/or from sea-ice melt in a region like the Labrador Sea, could induce a positive salinity bias at the surface in the models (which cannot be detected from the North Atlantic averaged profile), which would then be transported to the rest of the basin through the deep southward branch of the AMOC.

6.2.2 The Revelle factor

To explore whether the bias in the recent change in the North Atlantic CO_2 uptake in the CMIP5 models persists at the basin-scale or in localised areas, the Revelle factor in the models and real-world is explored. Indeed, since the Revelle factor γ describes the buffering capacity of the ocean to take up additional CO_2 (Chapter 1; where high Revelle factor values correspond to a low buffering capacity), models biases in γ would highlight key missing biogeochemical mechanisms in the concerned regions.

The observation-based Revelle factor (Figure 6.2a) was personally communicated by Siv K. Lauvset, one of the authors of the GLODAPv2 product (Lauvset et al., 2016), who calculated the Revelle factor on each binned grid cell before generating the basin-wide interpolation. The Revelle factor was calculated in 15 CMIP5 models (Table A.3) using in the Equation 1.14, the DIC and TA climatologies calculated over the periods 1996-2008 and 1972-2013, respectively (following discussions in Section 2.2.1).

While the CMIP5 models present on average no systematic error with the observation-based Revelle factor in the subtropical gyre (with a good model agreement characterised by a close to zero inter-model standard deviation; Figure 6.2c), they largely overestimate the observation-based values in the subpolar region (by about 5 units, especially in the Labrador Sea; Figure 6.2b). The overestimation of the Revelle factor in the CMIP5 models suggests that the models' subpolar surface waters have a lower ability to take up additional CO_2 from the atmosphere than in the real-world. As such, the systematic underestimation of the change in the North Atlantic air-sea CO_2 flux in the recent past by the CMIP5 models (identified in Chapter 5) might be linked to systematically misrepresented or missing mechanisms in the subpolar region.

To further explore the models' positive bias of the Revelle factor in the subpolar region, links are made to the subpolar seasonal cycle analysis conducted in Chapter 3. Indeed, most

of the CMIP5 models provided a poor representation of the seasonal $fCO_{2\text{-ocean}}$ signal in the subpolar region, except for the NorESM1-ME and GFDL-ESM2M models, which seemed to capture the spring bloom activity (c.f. Figure 3.16). We find that, for the subpolar Revelle factor, the NorESM1-ME and GFDL-ESM2M models also have the smallest mean absolute errors with the observation-based values among the others 13 CMIP5 models, with the respective area-weighted values of 0.8 and 1.2. The IPSL-CM5B-LR model, which was identified as the CMIP5 model with the poorest skill for the capture of the surface $fCO_{2\text{-ocean}}$ seasonal cycle (c.f. Figure 3.17), also provides for the subpolar Revelle factor study, the highest mean absolute error among of all the CMIP5 models, with an area-weighted value of 4.3. These results suggest that substantially misrepresented biological mechanisms in the subpolar gyre could explain the unrealistic CO_2 buffering capacity in that region. The subpolar region therefore appears as a crucial area for setting up the appropriate amount of CO_2 that the North Atlantic surface waters are able to take up from the atmosphere.

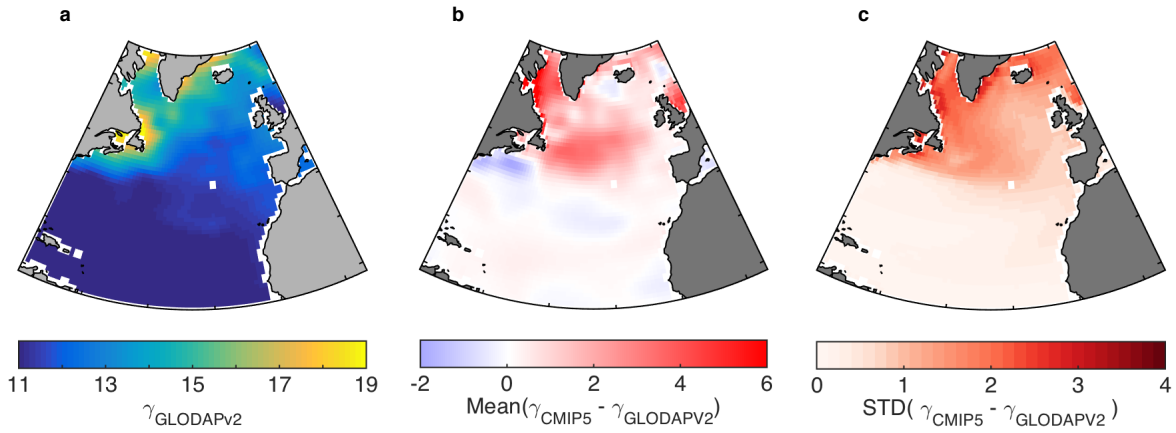

Figure 6.2: Model evaluation of the Revelle factor in the North Atlantic. **a**, Revelle factor γ based on the GLODAPv2 DIC and TA climatological binned values (Lauvset et al., 2016) (personal communication from S. K. Lauvset), **b**, Mean difference between the 15 CMIP5 models (Table A.3) and the observation-based Revelle factor, with the inter-model difference variability shown in **c**.

Although the exact reason behind the models' $fCO_{2\text{-ocean}}$ trend bias in the North Atlantic remains unidentified, the salinity profiles and the Revelle factor preliminary analyses suggest areas of further investigations, which should also be addressed within the modelling community in order to improve the simulation of the marine CO_2 system in ESMs.

6.3 Perspectives: extension of the investigation of the recent change in the oceanic CO_2 uptake to the global scale

The systematic underestimation of the recent change in the North Atlantic CO_2 sink by the CMIP5 models fundamentally questions their ability to predict the future CO_2 uptake in that basin. Nevertheless, the models' discrepancy with the observation-based estimate found in the North Atlantic may not persist in other regions, where the models might in fact provide a counter-balance effect, leading to a correctly simulated global CO_2 uptake. To evaluate the CMIP5 models' global CO_2 oceanic uptake, the change in the global sink must be first carefully determined in the real-world. As such, the interpolation technique developed in this thesis (Chapter 4) was extended to the global scale (Section 6.3.1), but the inconclusive results in poorly-sampled basins lead us to discuss zones of improvements in the interpolation technique itself (Section 6.3.2). Finally, a preliminary data assimilation study, which aimed to combine the strength of the observations and models to provide global $fCO_{2\text{-ocean}}$ estimates, is presented (Section 6.3.3).

6.3.1 Extending the MLR-based study to the globe

To test whether the overestimation of the $fCO_{2\text{-ocean}}$ trend is only specific to the North Atlantic basin, the interpolation approach used in this thesis was preliminary extended to other basins[1]. This preliminary work aimed to quantify robust uncertainties on the $fCO_{2\text{-ocean}}$ trend over the period 1992-2014 in all the ocean basins (and therefore at the global scale), and subsequently evaluate the CMIP5 models.

Initial results showed that in some areas (e.g. the Indian Ocean, the Southern Ocean), the interpolation uncertainty was either too wide to provide a meaningful model evaluation or the methods introduced a bias in the prediction of the time-varying surface $fCO_{2\text{-ocean}}$. The poor predictive skill of the MLR interpolation technique in those regions might be due to (1) the limited available observations in those basins that do not enable the capture of the ranges of the spatial and temporal variability, and (2) the explanatory variables (i.e. SST, MLD, atmospheric xCO_2 and position) being inadequate to fully describe the surface $fCO_{2\text{-ocean}}$ in those

[1] This work was based on the methods developed in Chapter 4 and was carried out with Tobia Tudino through the NERC Carbon Uptake Revisited - Biases Corrected using Ocean Observations (CURB CO^2) project, a project leaded by Paul Halloran.

areas (further improvements on the MLR will be discussed later in the chapter). However, in some other areas (e.g. North Pacific, South Atlantic), the uncertainties on the time-varying $fCO_{2\text{-ocean}}$ over the period 1992-2014 and on the resulting trends were robust, against which the CMIP5 models were evaluated. We found that in those well-sampled regions, the CMIP5 models also seemed to overestimate the surface $fCO_{2\text{-ocean}}$ trend estimate over the period of study (further sensitivity and internal variability analyses are required to confirm this result). While we cannot assess with the present interpolation technique and with the available CO_2 observations whether the bias in the CMIP5 models' $fCO_{2\text{-ocean}}$ trend persists at the global scale, evidence suggests that the model-observation discrepancy remains across major ocean basins. As such, the exact reason(s) behind the models' $fCO_{2\text{-ocean}}$ trend bias should be further investigated, at least in the North Atlantic, in order to provide appropriate improvements for the next generation of Earth System Models (ESMs), and potentially provide corrected future predictions of the oceanic CO_2 sink.

6.3.2 Improving the interpolation technique

The discussion has so far mainly been oriented on the biogeochemical biases in the CMIP5 models, yet improvements on the methods used to determine the real-world's global change in the CO_2 ocean sink should also be considered. Indeed as mentioned above, the MLR interpolation technique and the basin-wide uncertainty assessment developed in this thesis were not conclusive for some ocean basins (e.g. the Indian Ocean) over the period 1992-2014, leading to unquantified robust global estimates. To improve the $fCO_{2\text{-ocean}}$ predictions in those basins, further investigations could be made:

- The optimal latitudinal subdivision could be determined for each ocean basin. Indeed, the MLR method applied in this study used a 5° latitude band width in each basin (as for the North Atlantic), which may well be inappropriate in an area with sparse observations.

- The period of study (i.e. 1992-2014) could be reduced, especially if the basins are poorly sampled in the 1990s or early 2000s. A reduction of the period of study would potentially remove the bias introduced by the MLR on the annual $fCO_{2\text{-ocean}}$ time series, but would ultimately constrain the model evaluation to the comparison of different

phases of internal variability.

- The MLR itself could be improved by adding other explanatory variables, such as sea surface height, which is related to ocean circulation and wind patterns, or chlorophyll-a, which is a more direct proxy than the MLD to describe the biological influences on the surface $fCO_{2\text{-ocean}}$.

However, by extending the MLR with additional explanatory variables, three main issues might arise: (1) depending on the amount of information available within a latitude band width over the period of study, the MLR might overfit the data and would therefore provide poor predictive skills, (2) depending on the chosen additional explanatory variables, the period and locations at which all variables are available might limit the range of the study (this is particularly the case with chlorophyll-a data, which are usually available on relatively short timescales and are not available all through the year for all grid cells depending on the satellite position and cloud cover; for example, if the latest version of the OceanColour-CCI product was used in the MLR, our study would have started from late 1997 only, which would have impacted the trend analysis by including more internal variability in the system compared to a 23-year long study), and (3) the justification of the use of the uncertainties determined from the CMIP5-based MLR analyses on the observation-based MLR results (as achieved in Figure 4.9) might be compromised. Indeed, if we imagine that chlorophyll-a is an additional explanatory variable in the MLR interpolation technique, and consider that the CMIP5 models poorly represent the spring bloom (as discussed in Chapter 3), the CMIP5-based MLR would describe a less variable system and might therefore provide good predictive skill at the locations and times at which measurements were made. On the other hand, MLR based on the real-world, which does contain the spring bloom signal, might in fact poorly predict the spring bloom and return a poor predictive skill (the MLR would likely underestimate the true signal). As such, due to the poorly simulated biological activity in the models, the interpolation uncertainty determined from the CMIP5-based MLR analyses might underestimate the interpolation uncertainty in the real-world, and hence not be appropriate.

6.3.3 Developing a global data assimilation product

While the addition of further complexity in the interpolation technique would potentially allow us to better describe the variability in the surface $fCO_{2\text{-ocean}}$ in the real-world, we would also rely more on the CMIP5 models capturing that variability in order to provide an appropriate uncertainty assessment. To improve the simulation of the real-world variability by models, and still allow us to provide basin-wide $fCO_{2\text{-ocean}}$ observation-based estimates, the assimilation of surface $fCO_{2\text{-ocean}}$ observations in an ocean model offers a useful alternative to potentially improve our understanding of the global change in the oceanic CO_2 uptake. Two previous studies (Valsala and Maksyutov, 2010; While et al., 2012) assimilated global $pCO_{2\text{-ocean}}$ observations in a physical-biogeochemical coupled ocean model for the period 1996-2004 and for the year 2006, respectively; both studies highlighted the benefits of constraining the simulation with surface CO_2 observations (even in poorly sampled regions) to minimise the model's biases.

To provide a global $fCO_{2\text{-ocean}}$ product that could be of interest for both the modelling and observational communities, a new $fCO_{2\text{-ocean}}$ assimilation was generated, by extending the work of While et al. (2012). We assimilated daily $fCO_{2\text{-ocean}}$ data from SOCATv4 (Bakker et al., 2016) into the coupled NEMO-HadOCC ocean model (Palmer and Totterdell, 2001; Madec, 2008; Megann et al., 2014), forced with the ERA-Interim three hourly surface conditions (Dee et al., 2011). The data assimilation simulation started in 1990 from a control run, which was designed to begin to spin up the physical ocean state[2]. The initial aim was to produce the longest $fCO_{2\text{-ocean}}$ data assimilation record available in order to (1) describe various modes of variability within the surface CO_2 system, and (2) provide an additional comparative study to the North Atlantic $fCO_{2\text{-ocean}}$ trend over the period 1992-2014 (as well as to the other basins). As such, the data assimilation was set to produce basin-wide $fCO_{2\text{-ocean}}$ estimates at the monthly frequency from 1st January 1990 to 31st December 2014. However, due to persistent modelling errors (e.g. the model was crashing after the assimilation of relatively high $fCO_{2\text{-ocean}}$ values in coastal areas), the data assimilation had to

[2]To move the ocean physics towards equilibrium (at least in the upper part of the circulation), the control run was forced with the ERA-Interim surface conditions over two cycles of the full ERA-Interim product. The first-cycle run covered the period 1943-1978, using the atmospheric CO_2 mixing ratio of those years but using the available ERA-Interim conditions of 1979-2014. The second-cycle run covered the period 1979-2014, using the corresponding atmospheric CO_2 mixing ratio and surface conditions. The control run was initialised from an observation-based product at the beginning of the first-cycle run, in 1943. The data assimilation simulation was started from the year 1990 of the control run (second cycle), which therefore had by then 48 years of spinup.

be stopped in July 2004, until further investigations. Due to time constraints, the initial 14.5 years of assimilated data have not yet been analysed. While we would expect the data assimilation to provide an improved surface $fCO_{2\text{-ocean}}$ prediction compared to the control run (without data assimilation), we would have to be cautious on how the model adjusted the surface $fCO_{2\text{-ocean}}$ to the observations. For instance, the model could set the surface DIC to unrealistic values in order to return a surface $fCO_{2\text{-ocean}}$ that is close to the observations.

The improvement of the global MLR interpolation method with its associated basin-wide uncertainty, and the development of $fCO_{2\text{-ocean}}$ data assimilation on multi-decadal timescales would likely enhance our understanding of the recent change in the oceanic CO_2 sink, especially in areas of limited observations. However, such methods rely on observations and hence cannot directly inform us about the future change of the global CO_2 oceanic sink. In a final piece of preliminary analyses, the future of the North Atlantic and global change in the CO_2 oceanic are studied.

6.4 Perspectives: provide bias-free future predictions for the North Atlantic and global CO_2 oceanic sink

The main result found in this thesis (i.e. the systematic underestimation of the recent change in the North Atlantic CO_2 uptake in the CMIP5 models) questions the reliability of the latest generation of ESMs to predict the future contribution of the oceans in mitigating climate change. To initiate the discussion of bias-corrected future predictions of the oceanic CO_2 sink, an illustrative emergent constraint analysis (Section 6.4.1) and an on-going work based on a set of initialised ocean simulations (Section 6.4.2) are here presented.

6.4.1 Emergent constraint

To investigate the potential implications of the underestimation of the recent change North Atlantic CO_2 uptake by the CMIP5 models on the future evolution of that sink, an illustrative emergent constraint analysis is undertaken. Emergent constraints have been increasingly applied within climate science e.g. Cox et al. (2013) and more recently used in

Kwiatkowski et al. (2017), utilising relationships that emerge between an observable short term property and a longer term feature that cannot be observed (usually related to a future change) within models, to observationally constrain future projections.

Here, an emergent relationship is identified between the North Atlantic $fCO_{2\text{-ocean}}$ trends over the period 1992-2014 in the CMIP5 models and the cumulative CO_2 anomaly from 2000 to 2099 simulated by those models (dark grey line in Figure 6.3). The cumulative CO_2 anomaly in each of the 18 available CMIP5 models (Table A.3) was deduced by (1) calculating the annual air-sea CO_2 flux time series from 2000 to 2099 (using the business-as-usual scenario), (2) calculating each models' annual mean anomalies from its value in the year 2000, and (3) multiplying the averaged time series by the length of the period of study (from 2000 to 2099). The emergent linear relationship returns a R-squared value of 0.55 and the gradient and intercept p-values are significant at the 95% confidence level. Knowing the North Atlantic $fCO_{2\text{-ocean}}$ trend in the real-world over the period 1992-2014 with confident uncertainties (blue line and grey shading in Figure 6.3), we use the CMIP5-based emergent relationship with the corresponding prediction intervals (light grey lines in Figure 6.3) to infer the cumulative CO_2 anomaly that the real-world would take up over the 21st century (dashed lines in Figure 6.3). While the emergent relationship illustrates a possible future outcome for the North Atlantic CO_2 sink, it relies on an extrapolation (Figure 6.3), as opposed to an interpolation (as in e.g. Cox et al., 2013, which had the short term observed property falling within the corresponding modelled values). As such, the result presented here should be interpreted with caution[3]. The illustrative emergent constraint analysis suggests that, should the observed long-term trend prevail, the North Atlantic might take up an additional 32 to 61 PgC from the atmosphere over the 21st century, relative to the year 2000 (blue star values in Figure 6.3); a result on average about three times larger that predicted by the latest generation of climate models (i.e. 14 ± 9 PgC; Figure 6.3).

[3]Ultimately, a more reliable emergent constrain would have arisen if the observation-based estimate had fallen within the range of the models' $fCO_{2\text{-ocean}}$ trends, a configuration which would have however prevented the discussion around of the systematic model biases in Chapter 5.

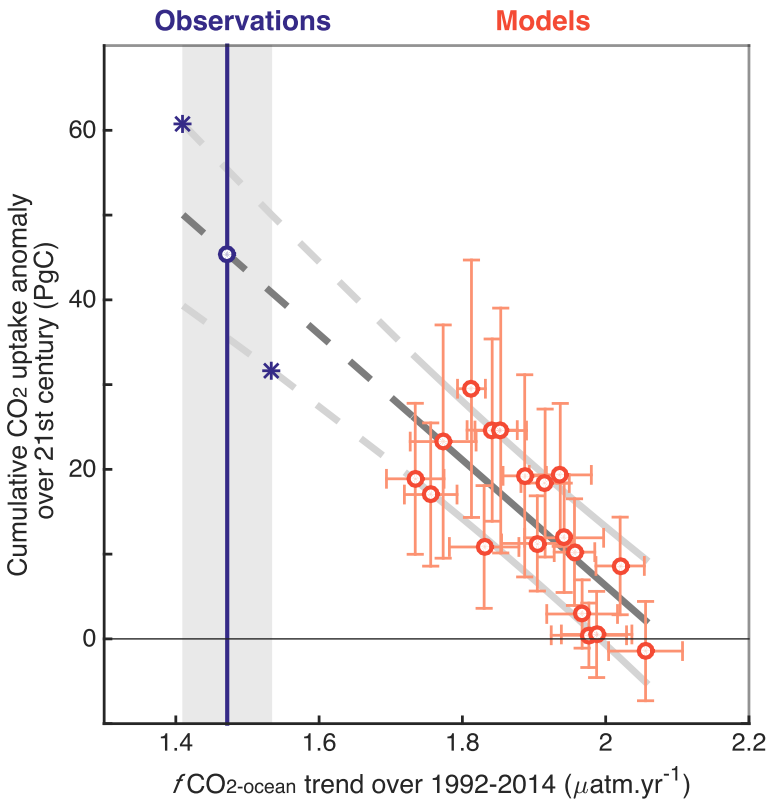

Figure 6.3: A potential future North Atlantic CO_2 uptake. Mean North Atlantic CO_2 uptake anomaly predicted for the period 2000-2099 relative to 2000 (blue circle), based on the extension of the linear relationship (dark grey line) between the $fCO_{2\text{-ocean}}$ trend over 1992-2014 and the cumulative CO_2 uptake anomaly in 18 CMIP5 models (orange circles, Table A.3), with the 1σ prediction interval (light grey lines). The horizontal and vertical error bars for each model correspond to the standard error on the linear trend and the 1σ on the cumulative CO_2 uptake anomaly, respectively. The grey shading is the 1σ uncertainty on the $fCO_{2\text{-ocean}}$ trend. The crosses cover the range of potential future CO_2 uptake anomaly in the North Atlantic (i.e. from 32 to 61 PgC).

While the emergent constraint approach allows us to quantify a possible range for the future North Atlantic CO_2 uptake, the emergent relationship is based on $fCO_{2\text{-ocean}}$ trends that are calculated over a period (i.e. 1992-2014) that might not be long enough to detect the forced signal from the unforced background (as suggested by McKinley et al., 2011, which identified a minimal period of 25 years to detect the anthropogenic forcing in the North Atlantic). As such, to provide more robust indicator for the long-term change and hence for the emergent constraint study, the MLR analysis and the CMIP5 models evaluation respectively carried out in Chapters 4 and 5 should be extended to a minimal 25-year long study, once new surface CO_2 measurements available. Such extension may allow us to verify that the overestimation of the CMIP5 models' $fCO_{2\text{-ocean}}$ trends persists on longer timescales

and therefore confirm that the models' bias does not result from a poorly simulated phase of natural variability that the real-world is potentially describing.

6.4.2 Bias-free ocean simulations

By reproducing the rate of increase in the recent North Atlantic CO_2 sink strength, the simulations conducted in Chapter 5 indicated that a coupled physical and biogeochemical ocean model initialised from observations can provide unbiased $fCO_{2\text{-ocean}}$ trends. Based on these results and assuming that the bias in the $fCO_{2\text{-ocean}}$ trends in the CMIP5 models persist in the future, we asked the question: Could a similar set of simulations provide a bias-corrected version of the CMIP5 models for the future predictions of the global (and North Atlantic) CO_2 oceanic sink? To answer this question, a set of twelve new simulations were run, principally over the 21st century[4]. While one of the five ocean simulations generated in Chapter 5 was forced with the reanalysis product ERA-Interim (Dee et al., 2011), the set of simulations forced in the future can only be achieved with the future surface conditions, which were taken from the CMIP5 archive (Table A.3). As such, the twelve simulations are divided into four groups, each of them being forced with the respective surface conditions[5] taken from the CanESM2, GFDL-ESM2M, HadGEM2-ES and IPSL-CM5A-LR models. For each of the four groups, three simulations were generated:

- The first simulation, the spinup phase, was forced from January 1950 to December 1999 with the surface conditions taken from the corresponding historical CMIP5 outputs. The aim of this simulation was to spinup the surface and intermediate ocean physics, rather than the biogeochemical fields, which were initialised in 2000 in the two following simulations.

- The second simulation was started from the end of the spinup simulation, and was initialised as in the simulations in Chapter 5, except that initialisation date was in January 2000 (instead of January 1979) and that the GLODAPv2 product (Lauvset et al., 2016) was used instead of GLODAPv1 (Key et al., 2004), because the former is more representative of the 2000s than the latter. The simulation was forced with the surface

[4]This work was supported by the NERC CURB CO^2 project.

[5]As a reminder, the surface conditions are defined by the surface air temperature, snowfall, specific humidity, vector winds, precipitation, downwelling shortwave and longwave radiation.

conditions of the corresponding CMIP5 model's outputs generated from the historical experiments from 2000 to 2005 and from the RCP8.5 scenario until the end of the 21st century. The prescribed atmospheric CO_2 mixing ratio followed the RCP8.5 scenario values. This simulation refers hereinafter as the "RCP8.5-forced simulation".

- The third simulation was set as the RCP8.5-forced simulation, except that it was forced with the surface conditions of the corresponding CMIP5 model's outputs generated from the RCP2.6 scenario (i.e. the climate change mitigation actions scenario) (Taylor et al., 2012), and the atmospheric CO_2 mixing ratio was set to the values returned by the RCP2.6 scenario. This simulation refers hereinafter as the "RCP2.6-forced simulation".

As discussed in Chapter 5 (c.f. Section 5.3.2), the quantification and removal of the model drift is crucial to provide appropriate studies of the time-varying behaviour of the $fCO_{2\text{-ocean}}$ system. As such, eight additional simulations with constant atmospheric CO_2 were performed. Specifically, the constant atmospheric CO_2 runs started in 2000 with the identical initialisation as the RCP8.5-forced and RCP2.6-forced simulations, ended in 2099, and were forced with the surface conditions produced by the four selected CMIP5 models (i.e. CanESM2, GFDL-ESM2M, HadGEM2-ES and IPSL-CM5A-LR) for the scenarios RCP8.5 and RCP2.6. Each of the RCP8.5-forced and RCP2.6-forced simulations therefore had a corresponding version with constant atmospheric CO_2 mixing ratio. However due to time constraints, the appropriate model-drift correction on long-term period has not yet been assessed. The results presented here have therefore not been model-drift corrected.

Initial results suggest that, under the RCP8.5 scenario, the CMIP5 models could underestimate by $\sim$40% the amplitude of the air-sea CO_2 flux at the end of the century compared to the flux simulated by the set of initialised runs within that period (red lines in Figure 6.4). However, under the RCP2.6 scenario, the distinction between the CMIP5 model simulations and the ocean-only simulations is less clear (blue lines in Figure 6.4). In the North Atlantic (not shown) similar results are found, with more inter-annual and pentadal variability than at the global scale. These results suggest that the biases in the DIC and TA in the CMIP5 models are particularly important in controlling the response to the intense increase in atmospheric CO_2 concentrations (described by the RCP8.5 scenario).

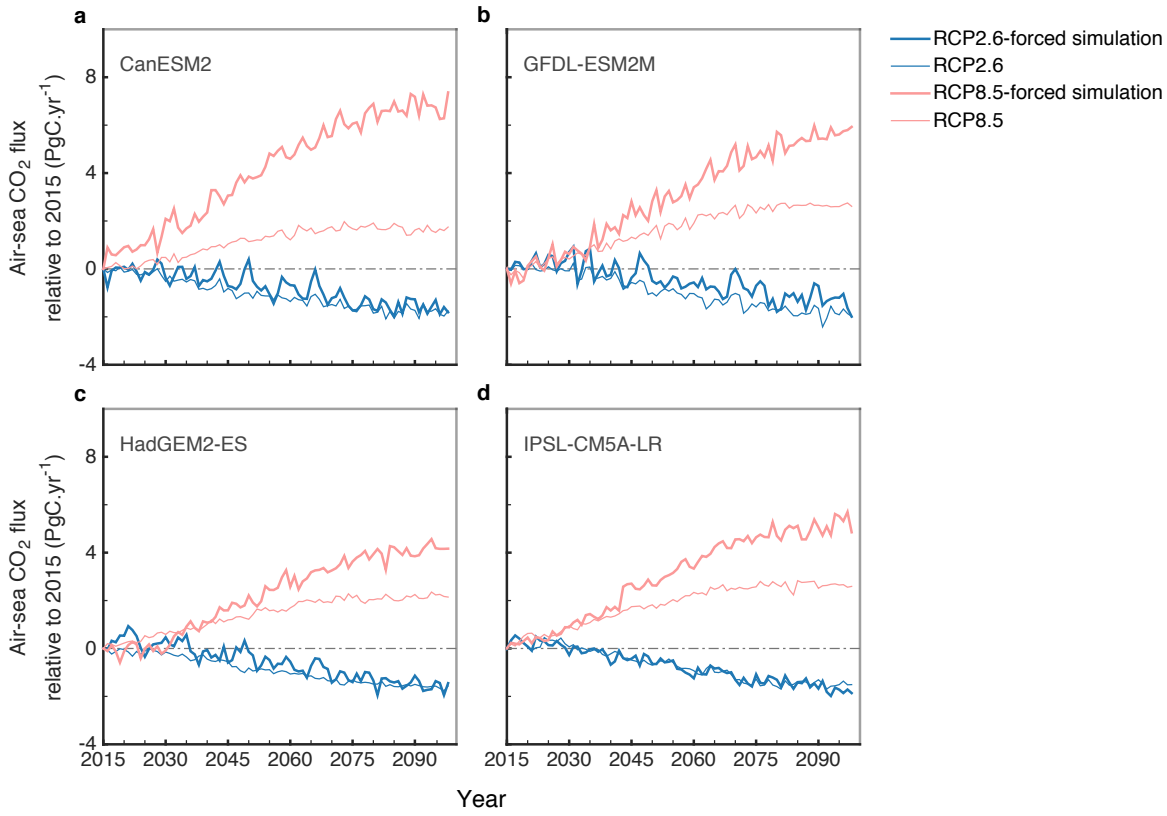

Figure 6.4: Bias-corrected future global oceanic CO_2 sink. Globally area-weighted annual means of the air-sea CO_2 flux produced by the RCP8.5-forced and RCP2.6-forced simulations (in thick red and thick blue lines, respectively) and by the CMIP5 models for the RCP8.5 and RCP2.6 scenarios (in thin red and thin blue lines, respectively. Panels **a**, **b**, **c** and **d**, correspond to the CanESM2, GFDL-ESM2M, HadGEM2-ES and IPSL-CM5A-LR surface forced conditions and to the studied CMIP5 models, respectively. All time series have been initialised to the the year 2015; as such, negative air-sea CO_2 flux values do not necessarily mean that the ocean is outgassing, but indicates that the flux has decreased since the 2015. No model-drift corrections have been here applied.

Overall, to understand, quantify and predict the variability of the CO_2 oceanic sink from inter-annual to multi-decadal timescales, models and observations should be used together, as the strength of models can help overcome the limitations of the observations and vice versa. To make further use of their complementarity, the scientific community should (1) extend the number and coverage of CO_2 measurements (e.g. within Southern Ocean), and (2) tackle the biogeochemical model biases (e.g. TA, surface $fCO_{2\text{-ocean}}$ seasonality and multi-decadal change), potentially through the improvement of the representation of surface biological processes and ocean circulation. While further analyses are required to provide robust time-varying global CO_2 estimates, the North Atlantic study undertaken in this thesis shows that by reconciling model and observation data, important progress can be made in our understanding of how models capture the recent change in the CO_2 uptake in that basin, and the real-world change can be determined with confidence.

Appendix A

Data processing

The data-processing of all products used throughout the thesis are developed in this appendix, starting with the CMIP5 model data (Section A.1) and following with all observational-based products (Section A.2).

A.1 The CMIP5 model data

Models description: Throughout the thesis, a total of 20 different CMIP5 models were used. The general characteristics of those models are described in Table A.1, and further details on their biogeochemical models are presented in Table A.2.

Model	Atmosphere	Ocean	Biogeochemistry	Reference
CESM1-BGC	**CAM4**	**POP2**	**BEC**	Gent et al., 2011
National Center for Atmospheric Research	1.25°x0.9°	60 levels		Long et al., 2011
	26 levels			
CMCC-CESM	**ECHAM5**	**OPA8.2**	**PELAGOS**	Vichi et al., 2011
Centro Euro-Mediterraneo per I Cambiamenti Climatici	3.75°x3.75°	ORCA2 (∼0.5-2°)		Vichi et al., 2007
	9 levels	31 levels		
CanESM2	**CGCM4/CanCM4**	**OGCM4/CanOM4**	**CMOC**	Zahariev et al., 2008
Canadian Centre for Climate Modelling and Analysis	35 levels	1.41°x0.94°		Chylek et al., 2011
		40 levels		
GFDL-ESM2G	**AM2**	**GOLD**	**TOPAZ2**	Dunne et al., 2012
Geophysical Fluid Dynamics Laboratory	2.5°x2.0°	0.3-1°		Dunne et al., 2013
	24 levels	63 levels		
GFDL-ESM2M	**AM2**	**MOM4p1**	**TOPAZ2**	Dunne et al., 2012
Geophysical Fluid Dynamics Laboratory	2.5°x2.0°	0.3-1°		Dunne et al., 2013
	24 levels	50 levels		
GISS-E2-H-CC	**GISS ModelE**	**HYCOM**	**NOBM**	Sun and Bleck, 2006
NASA Goddard Institute for Space Studies	2.5°x2°	1°x cos(*lat*)		Schmidt et al., 2014
	40 levels	26 levels		Romanou et al., 2014
GISS-E2-R-CC	**GISS ModelE**	**Russel**	**NOBM**	Hansen et al., 2007
NASA Goddard Institute for Space Studies	2.5°x2°	1°x1.25°		Schmidt et al., 2014
	40 levels	32 levels		Romanou et al., 2014
HadGEM2-CC	**MetUM**	**MetUM**	**diat-HadOCC**	Palmer and Totterdell, 2001
Met Office Hadley Centre	1.875°x1.25°	1°x0.3-1°		Collins et al., 2011
	60 levels	40 levels		
HadGEM2-ES	**MetUM**	**MetUM**	**diat-HadOCC**	Palmer and Totterdell, 2001
Met Office Hadley Centre	1.875°x1.25°	1°x0.3-1°		Collins et al., 2011
	38 levels	40 levels		
IPSL-CM5A-LR	**LMDZ5A**	**NEMOv3.2 OPA**	**PISCES**	Aumont and Bopp, 2006
Institut Pierre-Simon Laplace	1.875°x3.75°	ORCA2 (∼0.5-2°)		Dufresne et al., 2013
	39 levels	31 levels		
IPSL-CM5A-MR	**LMDZ5A**	**NEMOv3.2 OPA**	**PISCES**	Aumont and Bopp, 2006
Institut Pierre-Simon Laplace	2.5°x1.25°	ORCA2 0.5-2°		Dufresne et al., 2013
	39 levels	31 levels		
IPSL-CM5B-LR	**LMDZ5B**	**NEMOv3.2 OPA**	**PISCES**	Aumont and Bopp, 2006
Institut Pierre-Simon Laplace	1.875°x3.75°	ORCA2 0.5-2°		Dufresne et al., 2013
	39 levels	31 levels		
MIROC-ESM-CHEM	**MIROC-AGCM 2010+CHASER**	**COCO 3.4**	**NPZD**	Kawamiya et al., 2000
Japan Agency for Marine-Earth Science and Technology	2.8125°x2.8125°	1.4°x0.5-1.7°		Sudo et al., 2002
	80 levels	44 levels		Watanabe et al., 2011
MIROC-ESM	**MIROC-AGCM 2010**	**COCO 3.4**	**NPZD**	Kawamiya et al., 2000
Japan Agency for Marine-Earth Science and Technology	2.8125°x2.8125°	1.4°x0.5-1.7°		Watanabe et al., 2011
	80 levels	44 levels		
MPI-ESM-LR	**ECHAM6**	**MPIOM**	**HAMOCC5.2**	Giorgetta et al., 2013
Max Planck Institute for Meteorology	1.875°x1.875°	1.5°x1.5°		Ilyina et al., 2013
	47 levels	40 levels		Jungclaus et al., 2013
MPI-ESM-MR	**ECHAM6**	**MPIOM**	**HAMOCC5.2**	Giorgetta et al., 2013
Max Planck Institute for Meteorology	1.875°x1.875°	0.4°x0.4°		Ilyina et al., 2013
	95 levels	40 levels		Jungclaus et al., 2013
MRI-ESM1	**MRI-AGCM3**	**MRI.COM3**	**NPDZ**	Yukimoto et al., 2011
Meteorological Research Institute	1.125°x1.875°	1°x0.5°		Adachi et al., 2013
	48 levels	51 levels		
NorESM-ME	**CAM4-Oslo**	**MICOM**	**HAMOCC5.1**	Bentsen et al., 2013
Norwegian Climate Centre	2.5°x1.9°	1.125°		Tjiputra et al., 2013
	26 levels	53 levels		
bcc-csm1-1-m	**BCC_AGCM2.1**	**MOM4_L40v2**	**MOM4 FMS**	Wu et al., 2008
Beijing Climate Centre	1.125°x1.125°	1°x0.3-1°		Wu et al., 2013
	26 layers	40 levels		
bcc-csm1-1	**BCC_AGCM2.2**	**MOM4_L40v1**	**MOM4 FMS**	Wu et al., 2008
Beijing Climate Centre	2.8125°x2.8125°	1°x0.3-1°		Wu et al., 2013
	26 layers	40 levels		

Table A.1: List of the CMIP5 models main characteristics.

BGM	Nutrients	Phytoplankton	Zooplankton	Redfield
BEC	5 (NO_3,NH_4,PO_4,SiO_4,Fe)	3 (diatom, pico-nano phytoplankton, diazotrophs)	1	C:N:P
PELAGOS	5 (NO_3,NH_4,PO_4,SiO_4,Fe)	3 (diatoms, flagellates, picophytoplankton)	3 (micro-mesozooplankton, heterotrophic, nanoflagellates)	
CMOC	1 (that includes NO_3,NH_4, urea)	1	1	C:N
TOPAZ2	5 (NO_3,NH_4,PO_4,SiO_4,Fe)	3 (pico-nanophytoplankton, diatoms/eukaryotes, diazotrophic	1	C:N
NOBM	4 (NO_3,NH_4,SiO_4,Fe)	4 (diatoms, chlorophytes, cyanobacteria, coccolithophores)	1	C:N
diat-HadOCC	4 (NO_3,NH_4,SiO_4,Fe)	2 (diatom, non-diatom)	1	C:N
PISCES	5 (NO_3,NH_4,PO_4,SiO_4,Fe)	2 (nanophytoplankton, diatoms)	2 (micro-mesozooplankton)	C:N:P
HAMOCC5.2	3 (NO_3,PO_4,Fe)	1 (separated into diatoms and calcifiers)	1	C:N:P:Fe
HAMOCC5.1	4 (PO_4,NO_3,SiO_4,Fe)	1	1	P:C:N:ΔO_2

Table A.2: Main characteristics of the biogeochemical models used in the CMIP5 models. Complementary table version of Bopp et al. (2013).

Data download: Output from the CMIP5 models' (Taylor et al., 2012; Table A.1) was obtained from the Earth System Grid Federation (ESGF) online portal (https://esgf-node.llnl.gov/projects/cmip5/). From the available CMIP5 models (Taylor et al., 2012; Table A.3), the following fields were downloaded for the 1st ensemble member (labelled *r1i1p1* on ESGF) and for the historical experiment prior 2006[1] and the RCP8.5 scenario (Riahi et al., 2007) from 2006 (text in brackets indicates the variable name labelled on ESGF):

- **daily** near-surface specific humidity (*huss*, in kg.kg^{-1}), near-surface air temperature (*tas*, in K), surface longwave (*rlds*) and shortwave (*rsds*) downwelling radiations (in W.m^{-2}), eastward (*uas*) and northward (*vas*) near-surface winds (in m.s^{-1}), precipitation (*pr*) and snowfall (*prsn*) fluxes (in kg.m^{-2}.s^{-1}, equivalent to mm.s^{-1} for a density of 1,000 kg.m^{-3}).

- **monthly** sea surface partial pressure (*spco2*), temperature (*tos*) and pressure (*psl*), potential temperature (*thetao*), salinity (*so*), and CO_2 air-sea flux (*fgco2*).

- **yearly** DIC (*dissic*) and TA (*talk*).

All variables are surface fields (i.e. 3-dimensional fields, with time×lat×lon), except for potential temperature, salinity, DIC and TA which are also available at depth (i.e. 4-dimensional fields, with time×depth×lat×lon). Note that the DIC and TA fields were also

[1]While all models provided historical fields to at least December 2005 and started their RCP8.5 runs from January 2006, the HadGEM2-ES and HadGEM2-CC models had their historical runs ending in November 2005 and their RCP8.5 runs starting in December 2005. This was taken into account when merging historical and RCP8.5 datasets together.

available at monthly frequency but surface only, which would have limited study at depth (Section 2.2.2). Additionally, the 1st ensemble member of the control run for the monthly sea surface *spco2*, the RCP2.6 scenarios for the monthly *fgco2* and for all daily fields mentioned above were downloaded for specific specific analysis/illustration in the thesis.

	daily fields	**daily fields (RCP2.6)**	**spco2**	**dpco2**	**spco2 (picontrol)**	**tos**	**psl**	**thetao**	**so**	**fgco2**	**fgco2 (RCP2.6)**	**dissic**	**talk**
Units	see text	see text	Pa	µatm	Pa	K	Pa	K	psu	$kg(C).m^{-2}.s^{-1}$	$kg(C).m^{-2}.s^{-1}$	$mol.m^{-3}$	$mol.m^{-3}$
CESM1-BGC			✓		✓	✓	✓	✓	✓	✓		✓	✓
CMCC-CESM			✓			✓	✓	✓	✓				
CanESM2	✓	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓
GFDL-ESM2G			✓			✓	✓	✓	✓	✓	✓	✓	✓
GFDL-ESM2M	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
GISS-E2-H-CC			✓		✓	✓	✓	✓	✓	✓		✓	
GISS-E2-R-CC			✓		✓	✓	✓	✓	✓	✓		✓	
HadGEM2-CC			✓		✓	✓	✓	✓	✓	✓		✓	✓
HadGEM2-ES	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
IPSL-CM5A-LR	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓
IPSL-CM5A-MR				✓		✓	✓	✓	✓	✓	✓	✓	✓
IPSL-CM5B-LR				✓		✓	✓	✓	✓	✓		✓	✓
MIROC-ESM-CHEM			✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
MIROC-ESM			✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
MPI-ESM-LR			✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
MPI-ESM-MR			✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
MRI-ESM1			✓			✓	✓	✓	✓			✓	✓
NorESM1-ME			✓			✓	✓	✓	✓	✓	✓	✓	✓
bcc-csm1-1-m			✓		✓	✓	✓	✓	✓	✓	✓		
bcc-csm1-1			✓		✓	✓	✓	✓	✓	✓	✓		
Used in chapter(s)	Ch. 5	Ch. 6	Ch. 3 Ch. 4 Ch. 5	Ch. 3 Ch. 4 Ch. 5	Ch. 5	Ch. 3 Ch. 4 Ch. 5	Ch. 3 Ch. 4 Ch. 5	Ch. 4	Ch. 4 Ch. 6	Ch. 1 Ch. 5 Ch. 6	Ch. 1	Ch. 2 Ch. 5 Ch. 6	Ch. 2 Ch. 5 Ch. 6

Table A.3: List per model of the downloaded CMIP5 fields. All fields correspond to the historical and RCP8.5 experiments, except for the additional "daily fields", "spco2" and "fgco2" columns which display the RCP2.6 experiments (for "daily fields" and "fgco2") and the control run (picontrol) (for "spco2"). The last row indicates in which chapters of the thesis the corresponding fields have been used; directly or indirectly. Note that more CMIP5 models provided some of the indicated variables but were excluded as they did not provide the necessary fields for the core of the analysis in Chapter 4, or provided poorly represented carbon cycle, such as the INM-CM4.0 model (Volodin et al., 2010; Jones et al., 2013)

Conversion steps: To provide identical units across the models' outputs, and similar units/variables as observational products, and hence facilitate inter-model and model-observation comparisons, few adjustments were made:

- The IPSL models (i.e. IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR) did not provide on the ESGF portal the variable *spco2* but provided *dpco2* in µatm, from which *spco2* can be deduced. To do so, *dpco2* was first converted into Pa by multiplying by 0.101325 (i.e. 1 atm = 101325 Pa). The partial pressure of CO_2 in the atmosphere was then calculated from the atmospheric xCO_2 forcing data (Riahi et al., 2007) and the respective modelled sea surface temperature and pressure fields (Equations 1.2; 1.3)

(Taylor et al., 2012; Pierrot et al., 2009) and finally subtracted by *dpco2*. The deduced *spco2* was therefore in Pa, as the rest of the models.

- To match with the SOCAT observational product, all models' surface *spco2* was converted into fugacity and from Pa to µatm, using the modelled sea surface temperature and pressure fields (Equations 1.1; Weiss, 1974) and by dividing by 0.101325. Note that for simplicity, and due to the different ways results from different experiments have been used, this conversion applied to the historical and RCP85 experiments but not to the picontrol experiment.

- The CO_2 air-sea flux variables was stored in kg(C).m^{-2}.s^{-1}, except for the CanESM2 model which was in kg(CO_2).m^{-2}.s^{-1} and was therefore multiplied by 12/44 to convert mass of CO_2 into C.

- Non-realistic salinity values for the CESM1-BGC model indicated that the units were in kg/kg, instead of psu (i.e. g/kg). The salinity field for this model was therefore multiplied by 1,000.

- The land mask for the *thetao* and *so* fields in the MRI-ESM1 model was set with zeros instead of "NaN". To provide clear land mask and avoid unrealistic MLD values at those grid points, all zero points were set to "NaN".

Regridding: The variety of model resolution and vertical/horizontal type of grids (Table A.1) lead to difficulties when comparing models together and/or with observational products that are generally based on regular grids. For example, the MPI-ESM-LR model uses for the ocean component a bipolar grid, whose poles are located in Greenland and on the coastal area of the Weddell Sea, while the medium resolution version (MPI-ESM-MR) uses a tripolar grid (Giorgetta et al., 2013). Within a model, the resolution also commonly varies. For example, the HadGEM2 ocean component provides a 1° meridional resolution from the pole to 30° and a slowing increasing resolution to 1/3° at the equator. To allow direct fields comparison, one crucial data processing step was therefore to regrid all model outputs into a regular grid.

All monthly (except *thetao* and *so*) and yearly fields indicated above were horizontally regridded to a transitional 180×360 grid using the bilinear interpolation within the CDO

package (http://www.mpimet.mpg.de/cdo) and were then adjusted to the SOCATv4 grid (also a $1° \times 1°$ grid, but with longitude going from -179.5°E to 179.5°E) using the bilinear interpolation function within the Python Iris package (http://scitools.org.uk/iris/docs/v1.7/index.html). The 4-dimensional *dissic* and *talk* fields were also vertically regridded according to the GLODAPv2 depth levels (Lauvset et al., 2016) using the CDO package (http://www.mpimet.mpg.de/cdo). However, since the first depth level in GLODAPv2 is 0 m and in the models is generally shallower (e.g. 5 m), the interpolation could not be computed between 0 m and the first level in the models (except for the NorESM1-ME model whose first depth level is 0 m). As such, the regridding provided "NaN" values at the surface and therefore led surface analyses of DIC and TA to be achieved at 10 m depth.

All daily surface fields, which were used to force an ocean model (Chapters 5 and 6), were temporally and spatially adjusted to models' set-up and grid. First, the time component of all variables was adapted to Gregorian calendars. Models that had a 365-day calendar and hence no leap years (GFDL-ESM2M, CanESM2 and IPSL-CM5A-LR) were added the extra day of February the 29th on leap years by repeating the field of February 28th. Model that had a 360-day calendar and hence 12 months of each 30-day long (HadGEM2-ES) were added an extra day at the end of the months of January, July, August, October and November on non-leap years and also March on leap years, by repeating the field of day 30. Then, all fields were spatially regridded to a transitional 180×360 standard-grid using the CDO package (http://www.mpimet.mpg.de/cdo), and then regridded to ORCA-1 (i.e. the grid of NEMO ocean model) using the Surface Interpolation Environment SOSIE (https://github.com/brodeau/sosie).

Mixed Layer Depth calculations: For an analysis conducted in Chapter 4, the MLD was required for each CMIP5 model. While the ESGF portal provided the variable *mlotst* (ocean mixed layer thickness defined by sigma-t) for few models, it was preferred, for clarity and uniformity reasons, to calculate the MLD for all models using the same method as in the observational-based product (Kara et al., 2000). Nevertheless, determining the MLD at each grid cell, for each month and for each model was a computationally expensive task. As such, calculations were achieved for the period 1990-2014, which is similar to the period of avail-

able MLD data from the observational-based product (from 1992; Table A.5), and using the first 40 depth levels of each model, if available. Indeed, each model has its specific vertical resolution, going from 31 to 70 depth levels (Table A.4). While the depth level 40 corresponds to relatively different depths across the models (from 985 m to 5720 m; Table A.4), the MLD is expected to be found somewhere between the surface and depth level 40, as the observed global MLD mean is 89 $\pm$ 268 m (Menemenlis et al., 2008). For each model, the MLD was determined as follows:

1. To facilitate the calculations steps, the salinity and potential temperature fields were horizontally regridded into a $1^\circ \times 1^\circ$ latitude$\times$longitude grid (not the SOCATv4 grid).

2. Salinity and potential temperature (in that order) were respectively converted into Absolute Salinity (AS) and Conductive Temperature (CT) using the Gibbs Seawater (GSW) Oceanographic Matlab toolbox v3.04 (http://www.TEOS-10.org; McDougall and Barker, 2011).

3. Density was calculated from AS and CT, also using the GSW toolbox.

4. MLD was calculated using CT and AS at the reference level (i.e. first depth level), from which the density criterion corresponding to a temperature change of 0.8°C was determined, and also using the density profiles (Kara et al., 2000). The method first determines the bottom of the uniform layer and then identifies the depth interval which contains the density criterion (this interval typically corresponds to the zone of transition between the well-mixed surface and intermediate waters). If this depth interval is found, the MLD therefore lies within that interval and is determined by linear interpolation (Kara et al., 2000). If the transitioning layer is not found, the steps are repeated using instead the second depth level as the reference level (Kara et al., 2000). If still not found, the MLD is set as NaN for this grid-cell.

5. The 3-dimensional MLD field was regridded to the SOCATv4 grid using the Python Iris package. The resulting calculations gave overall realistic values (Table A.4).

Model	Vertical resolution	Depth at level 40 or last available (m)	Mean MLD (m)
CESM1-BGC	60	985	78 ± 67
CMCC-CESM	31	5250	82 ± 123
CanESM2	40	5233	83 ± 93
GFDL-ESM2G	50	2049	80 ± 111
GFDL-ESM2M	50	2049	84 ± 119
GISS-E2-H-CC	33	5500	78 ± 121
GISS-E2-R-CC	32	4887	92 ± 170
HadGEM2-CC	40	5328	78 ± 73
HadGEM2-ES	40	5328	78 ± 71
IPSL-CM5A-LR	31	5250	77 ± 102
IPSL-CM5A-MR	31	5250	78 ± 102
IPSL-CM5B-LR	31	5250	76 ± 91
MIROC-ESM-CHEM	44	4525	90 ± 99
MIROC-ESM	44	4525	91 ± 97
MPI-ESM-LR	40	5720	84 ± 111
MPI-ESM-MR	40	5720	81 ± 108
MRI-ESM1	51	2500	66 ± 94
NorESM1-ME	70	1150	125 ± 112
bcc-csm1-1-m	40	5334	99 ± 178
bcc-csm1-1	40	5334	90 ± 126

Table A.4: Mixed Layer Depth processing. **2nd column**, Total number of depth levels within each model. **3rd column**, Approximated depth value at level 40 or at the last available level. **4th column**, area-weighted mean of the global MLD from 1992-2014 (same period as the used observational-based product), with standard deviation from the area-weighted mean.

Atmospheric xCO_2: For an analysis conducted in Chapter 4, a monthly 3-dimensional grid (time$\times$latitude$\times$longitude) storing annual values of xCO_2 was created. For each year, the same annual xCO_2 value was applied to each month, at each $1^\circ \times 1^\circ$ grid cell (following the SOCATv4 grid). Those xCO_2 values, obtained from the RCP Database Version 2.0.5 (Riahi et al., 2007), contain historical values up to 2005 and values following the RCP8.5 scenario from 2006.

A.2 Observational-based products

Here, the term "observational-based products" cover a wide range of datasets that use observations that have been processed in different manners: high-frequency measurements that have been merged into a regular spatial and temporal grid, localised measurements have

been interpolated using various statistical tools, data derived from satellite measurements or from reanalysis methods. The general information on all observational-based products used throughout the thesis, and the various data-processing steps are described in Table A.5.

Variable	Product	Description	Processing	Reference	Chapters
$fCO_{2-ocean}$	SOCATv4	1°×1° Monthly 1970-2015		Bakker et al., 2016 Sabine et al., 2013	Ch. 3, 4, 5
	Neural-Network	1°×1° Monthly 1982-2011		Landschützer et al., 2015 (personal com.)	Ch. 4
	Gap-filling	2.5°×2.5° Monthly 1985-2011		Jones et al., 2015	Ch. 4
$pCO_{2-ocean}$	Interpolation method	1°×1° Monthly 1990-2015		Iida et al., 2015	Ch. 4
DIC, TA	GLODAPv1.1	1°×1° 33 depth levels Climatology		Key et al., 2004	Ch. 2, 5
	GLODAPv2	1°×1° 33 depth levels Climatology	Converted into mol.m^{-3} by $\times 1025 \times 10^{-6}$	Lauvset et al., 2016	Ch. 1, 2, 5, 6
Salinity	GLODAPv2	1°×1° 33 depth levels Climatology		Lauvset et al., 2016	Ch. 6
SST	OISST v2	1°x1° Monthly 1981-present		Reynolds et al., 2007	Ch. 4, 5
MLD	ECCO2	0.25°×0.25° Daily 1992-present	Monthly averaged Interpolated into SOCATv4 grid	Menemenlis et al., 2008	Ch. 4
xCO_2	GLOBALVIEW-CO_2	Function of latitude 8-day frequency 1979-2014	Interpolated into daily Averaged into monthly Gridded to SOCATv4 grid	GLOBALVIEW-CO2, 2013	Ch. 4
SLP	NCEP/NCAR Reanalysis 1	2.5°×2.5° Monthly 1948-present	Interpolated to SOCATv4 grid Averaged into monthly Adjusted to SOCATv4 grid	Kalnay et al., 1996	Ch. 4
Wind Speed W	NCEP/NCAR Reanalysis 1	0.25°×0.25° 6-hourly 1948-present 0.995σ level	Calculated W^2 W^2 monthly average Adjusted to SOCATv4 grid	Kalnay et al., 1996	Ch. 4, 5
Bathymetry	ETOPO1	1' arc resolution	Adjusted to SOCATv4 grid Adjusted to "gap-filling" grid	Amante and Eakins, 2015	Ch. 3, 4, 5
Surface fluxes	ERA-Interim	0.75°×0.75° Daily 1979-2015	Adjusted to ORCA1 grid	Dee et al., 2011	Ch. 5

Table A.5: Description and processing steps of observational-based products. The term "surface fluxes" corresponds to: temperature at 2 m, specific humidity at 2 m, surface shortwave and longwave radiation downwards, total precipitation, snowfall, x-direction and y-direction wind components at 10 m.

Appendix B

The MSM060 voyage: a case study of the challenges associated with DIC and TA measurements

During this PhD, I had the opportunity to work as a carbon team member during two research voyages (DY040 and MSM060), from which I was able to particularly learn about voyage preparations, DIC and TA measurements and their Quality Control (QC) analysis. Since I led the complete QC analysis for the MSM060 carbon data only (I was involved in the initial QC steps for DY040), I therefore exclusively use the data from this voyage to highlight the different stages of providing high-quality biogeochemical measurements. Compared to the summarised analyses provided to the MSM060 voyage report and publication (Kartensen et al., *in prep.*), I deliberately include here expanded information and discussion about the various challenges faced and decisions we had to make during the measurement and QC phases. As such, this Appendix will provide a more comprehensive record of the data collection and processing than will be available in the peer-reviewed literature.

Note that the context and objectives of the MSM060 voyage, as well as technical carbon-related information, will be found in the voyage report or in Kartensen et al., (*in prep.*); which for synthesis purposes are not included here. Nevertheless, I will mention that the DIC and TA data were produced by the on-board carbon team (S. D. Jones, A. D. Lebehot, L. Cotrim Da Cunha and A. Carvalho) and supervised by the Principal Investigator U. Schuster (University of Exeter, UK); and that two Versatile INstruments for the Determination of

Titration Alkalinity (VINDTA) (Marianda, Kiel, Germany, serial numbers #064 and #065) were used for the measurements.

B.1 Carbon sampling and analysis strategies

Discrete seawater samples for DIC and TA measurements were collected from a rosette of 22 Niskin bottles. Specifically, seawater was sampled in 250 ml bottles, except for the top, bottom and two random Niskin bottles for which seawater was collected in 500 ml bottles to allow duplicate analysis and therefore conduct the precision analysis. Due to time constraints, we first decided to only analyse samples co-located with changes in the water mass distribution, and also selected depths between stations to optimise spatial interpolation throughout the whole section. The bottom and surface samples were always included in this subset selection of analysed bottles. The remaining samples were kept aside for potential further analysis. For each station, once the subset samples were analysed and first quality controlled, important gaps in the DIC and TA profiles or potential interesting features or outliers could be identified. The remaining unanalysed samples that would fill those gaps and hence potentially improve the continuity of the DIC and TA profiles, were analysed later in the voyage when time permitted. All remaining samples were discarded. In the end, about 75% of all samples were analysed.

Such sampling and analysis strategies are however not necessarily a standard approach, but was elaborated by the on-board team based on past-experiences. We specifically decided to sample all Niskins in order to produce relatively smooth profiles particularly near the significant changes in DIC and TA throughout the water column, and to facilitate and improve our decision-making regarding potential suspicious data during the first QC phase.

B.2 First quality control

Initial first QC was carried out on-board allowing any technical or instrumental issues to be addressed and solved immediately. The complete first QC was achieved post-voyage, once the quality controlled temperature and salinity data provided. Data impacted by technical errors (e.g. errors in liquid level sensors) were marked as “Bad” (i.e. flag 4), while suspicious

data (e.g. outliers within a vertical profile or in comparison of neighbouring profiles) were marked as "Questionable" (i.e. flag 3). The remaining data were flagged as "Acceptable" (i.e. flag 2). Nevertheless, the flag assessment was particularly challenging for most of the DIC measurements, due to some instrumental instabilities, which are described here:

- The DIC measurement is based on a titration by coulometry, a method that returns a cumulative counts at each incremented time step (SOP 2, Dickson et al., 2007). When the count difference between the current and previous time step (i.e. increment) is below a certain threshold (here set at 200), the titration has reached an end point. After reaching a minimum of four end points, the titration stops and the DIC value is deduced. However, repetitive communication issues between the coulometers and the controlling computers (particularly the one connected to VINDTA #065) led to the measure of unrealistic zero-count values, which postponed by few time steps the detection of a valid end point and consequently perturbed the final DIC value. Nevertheless in most cases, *true* DIC values could be retrieved; and by *true* we mean the count value that would have been selected for the final DIC calculation if those communication issues had not occurred. The retrieval of the true count values is explained with the following example.

  The zero-count value recorded at time step 5 leads to an unrealistic increment value and therefore impacts the increment at step 6 (Figure B.1a). However, since the count difference between steps 6 and 4 (i.e. 192) is below the set threshold (i.e. 200 counts) and that counts through the titration must be strictly increasing, both steps 5 and 6 would have triggered an end point in the absence of the communication issue. The first end point would have been at step 5 and the last one therefore at step 8. By applying this shift in the end points, the true count value (i.e. 190,639 at time step 8) can be retrieved and used to correct the default DIC calculations (i.e. using count value from step 10).

  However, depending on the time step in the titration at which the zero-count issue occurred, true count values could sometimes not be retrieved. In another example (Figure B.1b) and based on the above explanations, the true count value would have been reached at time step 8, which is also the time step when the zero-count issue occurred. In such configurations, the next count value (here 190319 at step 9) is used to calculate DIC. Although this correction does not provide the true count value, it improves the

count estimation compared to the value (i.e. at step 10) that would have been chosen otherwise. As such, we decided to flag as "Acceptable" DIC measurements based upon those zero-count corrections.

Note that the cause of the communication issues remains to this day unknown. However we do know that this problem persisted during the following voyage, suggesting that it was not a ill-installation in the setting up but rather an instrumental problem.

a

Time	Counts	Increment	End Point
1	95593	95593	0
2	180027	84434	0
3	189415	9388	0
4	190287	872	0
5	0	4294777015	0
6	190479	190473	0
7	190576	97	1
8	190639	63	2 => 4
9	190717	78	3
10	190831	114	4

b

Time	Counts	Increment	End Point
1	89795	89795	0
2	177215	87420	0
3	188785	11570	0
4	190026	1241	0
5	190090	64	1
6	190123	33	2
7	190212	89	3
8	0	4294777084	3 => 4
9	190319	190319	3
10	190397	78	4

Figure B.1: Examples of zero-count correction. Examples for which the true count value **a**, was retrieved, and **b**, could not be retrieved but still provides an improvement on the final count value used.

- While no obvious technical issues were found, it occasionally occurred that the DIC measurements were unstable, particularly on VINDTA #064. To identify whether those instabilities were persistent, test samples (i.e. mixture of random seawater) would be analysed for DIC only using the "junk" setting provided by the VINDTA software. The "junk" setting measures DIC similarly as the combined DIC-TA setting, with the exception that it does not execute the clear check, which tests the initial background noise of the coulometer. Since the reproducibility of count values through the junks was within our acceptable range, we therefore run a junk before each sample analysis and used the junk value instead of the DIC value returned by the DIC-TA setting, until stability in the measurements was reached again. Similarly, if obvious technical reasons affected the DIC measurements and if enough seawater remained in the sample, a junk routine was run and its result replaced the original DIC value. Replacements that matched the rest of the profile and results from stations in between were flagged as "Acceptable".

B.3 Precision analysis

The precision (i.e. the reproducibility of the experimental procedure) of the DIC and TA measurements is determined through the comparison of duplicate results within a 500 ml bottle (Dickson et al., 2007). Specifically, the precision is defined by the standard deviation across the duplicate differences, using the "Acceptable" data only. Since the precision should be independent from the seawater properties, we did not expect any distinctions between the duplicate analysis based on the 500 ml sample bottles and the one based on the (500 ml) CRM bottles. However, differences between the sample and CRM precision analyses were found, and as such the two precision analyses are presented and the potential cause(s) for their differences are then discussed.

Duplicate analyses

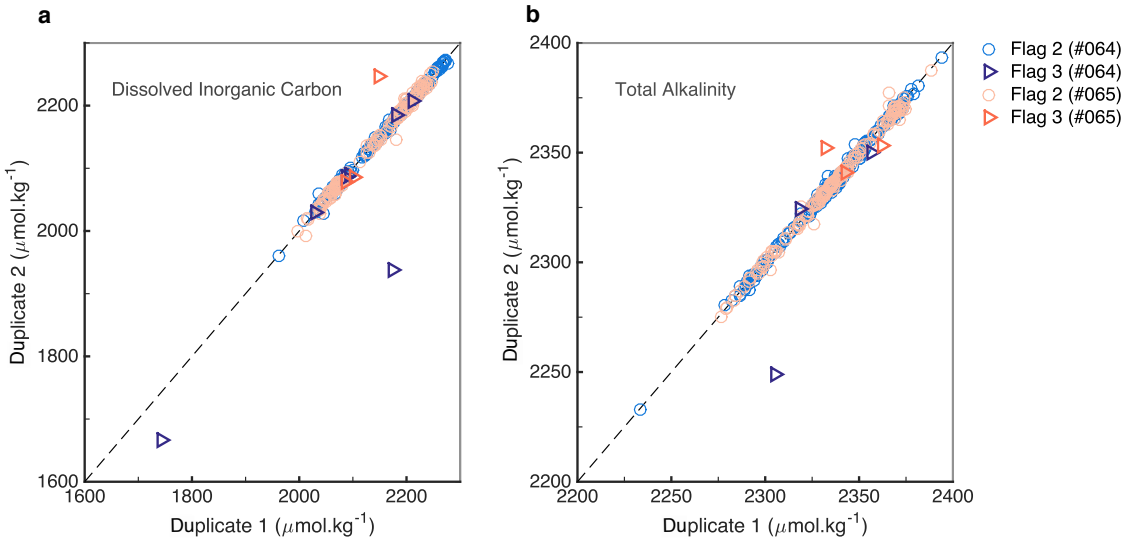

Figure B.2: Precision analysis on the sample duplicates. **a**, DIC, **b** TA, values for first duplicate versus the second duplicate, where blue and pink symbols correspond to measurements conducted on VINDTA #064 and #065, respectively. Circles correspond to flag 2 ("Acceptable") data, while triangles to flag 3 ("Questionable") data. The dashed line corresponds to the 1-to-1 line.

The sample duplicate analysis was achieved by using the 500 ml seawater samples that were systematically collected at the top and bottom Niskin bottles and two random depths in between; if chosen to be analysed. Duplicates that have been flagged as "Questionable" mostly show a lack of reproducibility (i.e. away from the 1-to-1 line) (Figure B.2), which adds confidence to our decisions made during the first QC steps. A focused study on the

"Acceptable" duplicates only shows that (1) the precision is better (i.e. the standard deviation across the duplicate differences is smaller) for the TA measurements than for the DIC measurements, and (2) the precision on VINDTA #064 is better than on #065 (Table B.1). The TA precision is consistent with previous precision analysis conducted on four different voyages (Schuster et al., 2014), while the DIC precision is on average 2.8 times higher than their study. Note however that Schuster et al. (2014) used different VINDTAs (SN #004 and SN #007) and small instrumental differences, which can lead to slightly precision difference and hence alter the comparison made here.

While CRMs are used to quantify the accuracy of a measurement and calibrate accordingly, they also provide information about the reproducibility of the measurements, by analysing the duplicates within a 500 ml CRM bottle. Similar analysis conducted on the sample duplicates is therefore reproduced here for the CRM duplicates (Table B.1). Precision on DIC and TA is similar across both VINDTAs and to the precision results found in Schuster et al. (2014).

	Sample duplicates			**CRM duplicates**		
	#064	#065	Both	#064	#065	Both
DIC (μmol·kg^{-1})	± 3.9	± 5.9	± 4.8	± 1.9	± 1.6	± 1.8
	(186)	(130)	(316)	(78)	(71)	(149)
TA (μmol·kg^{-1})	± 1.5	± 1.8	± 1.6	± 1.8	± 1.9	± 1.8
	(219)	(169)	(388)	(79)	(71)	(150)

Table B.1: Precision of DIC and TA measurements using the sample and CRM duplicates. Standard deviation of the differences between duplicate 1 and 2 for each VINDTA instrument and when combining both (i.e. column "Both"), calculated from the "Acceptable" duplicates. Values in bracket indicate the number of duplicates that were used to compute the statistics.

Discussion

Interestingly, VINDTA #065 presents different DIC precisions depending on whether sample or CRM duplicates are used, with a respective precision of 5.9 μmol·kg^{-1} and 1.6 μmol·kg^{-1} (Table B.1). Such discrepancy could have been due to the coulometer communication issues particularly persistent on VINDTA #065. As previously explained, some configurations in the communication issues led to non-retrieval of true DIC values, which could explain some of the noise in the precision. However, the ratio of affected measurements used in the sample

duplicate analysis (11%) and the ratio in the CRM duplicate analysis (9%) are similar. This therefore suggests that the difference in the DIC precision between sample and the CRM duplicate analyses is not particularly due to the coulometer communication issues.

The difference between the two precision analyses could also be due to the different numbers of duplicates that are used to compute the statistics. For instance, the DIC precision study on VINDTA #065 uses 130 duplicates for the sample duplicates analysis and 71 duplicates for the CRM duplicate analysis (Table B.1). To test the impact of the number of duplicates on the precision, a sensitivity analysis was run. The sensitivity analysis repetitively calculated the precision using a random subset of sample duplicates, whose size equalled the number of CRM duplicates. The sensitivity analysis concluded that the precision using the subsets sample duplicates is not significantly different from the precision using all sample duplicates (not shown). As such, the differences between the precision analyses run on the sample and on the CRM duplicates are not explained by the differences in the statistics. The reasons for the difference between the sample and the CRM duplicates precision analyses therefore remain unclear.

To investigate the discrepancy between the sample and CRM precisions, more laboratory work would be required. For example, one could switch the coulometer used on VINDTA #064 to the VINDTA #065 and vice-versa, and see if this device switch impact the precision analysis. This would provide an indication on whether the relatively low precision for DIC currently found for VINDTA #065 is due to coulometer itself or due to instrumental settings on the VINDTAs.

B.4 Calibration

The repetitive measure of CRMs (batch 159 from A. Dickson, Scripps Institution of Oceanography, USA) throughout the voyage allowed the quantification and correction of any potential instrumental offset and drift. All DIC and TA measurements were calibrated to the nearest-in-time analysed CRM.

Surprisingly, the offset in DIC and TA on VINDTA #065 is on average about 210 µmol·kg^{-1} and 270 µmol·kg^{-1}, respectively (Figure B.3). VINDTA #064 also presents a major offset for the DIC measurements with an overestimation of about 200 µmol kg^{-1} compared to the

batch value, while the offset for the TA measurements is about 30 µmol·kg^{-1}(Figure B.3). Such offsets indicate that some input parameters on the VINDTAs must have shifted from their initial calibration.

The identification of the calibration shift(s) and the reasons for such differences across the machines for the TA measurements have not been yet fully investigated. However software settings were checked, particularly to spot any differences between VINDTA #064 and #065 for the TA measurements. The only difference between the two machines found so far was a little, but maybe significant, difference in the TA titration curve around the location where the TA value is determined. To exactly identify the causations of the offset in the measurements and rectify accordingly, extensive laboratory work would be expected (e.g. test if the calibrations of the temperature sensors, pH probe, electrodes, pipettes volume, Titrino volume have shifted). Another batch of CRMs would also be required to verify the potential accuracy improvements, leading to significant additional laboratory costs.

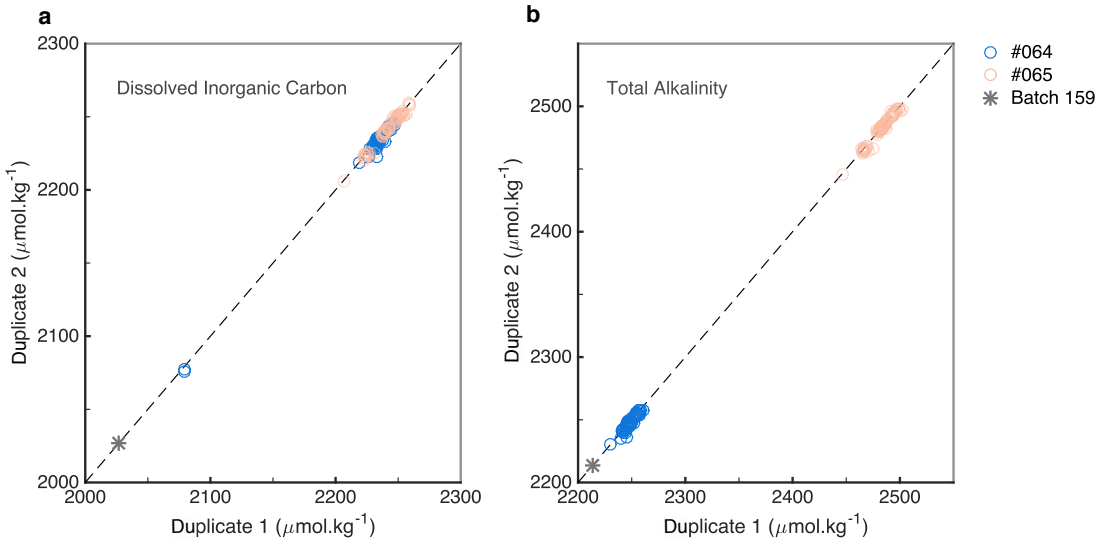

Figure B.3: Precision and accuracy analyses on the CRM duplicates. **a**, DIC, **b** TA, values for first duplicate versus the second duplicate, where blue and pink circles correspond to measurements conducted on VINDTA #064 and #065, respectively. The grey crosses correspond to the true CRM values from batch 159. The dashed line corresponds to the 1-to-1 line.

The large instrumental offsets question the assumption of linearity in the carbonate system made by CRM calibration. For example, if the instrument has an offset in the DIC of 210 µmol·kg^{-1} compared to the latest CRM measurement, that offset will be subtracted to the samples that are measured in between. To check that the linear assumption is here justifiable,

we present three DIC and TA profiles, in which most of the samples were measured on one machine and some additional samples on the other (Figure B.4). The few DIC and TA measurements made on one machine perfectly fit the rest of the profile measured on the other. This result therefore suggests that while the instruments are largely inaccurate, the CRM calibration fixes this inaccuracy.

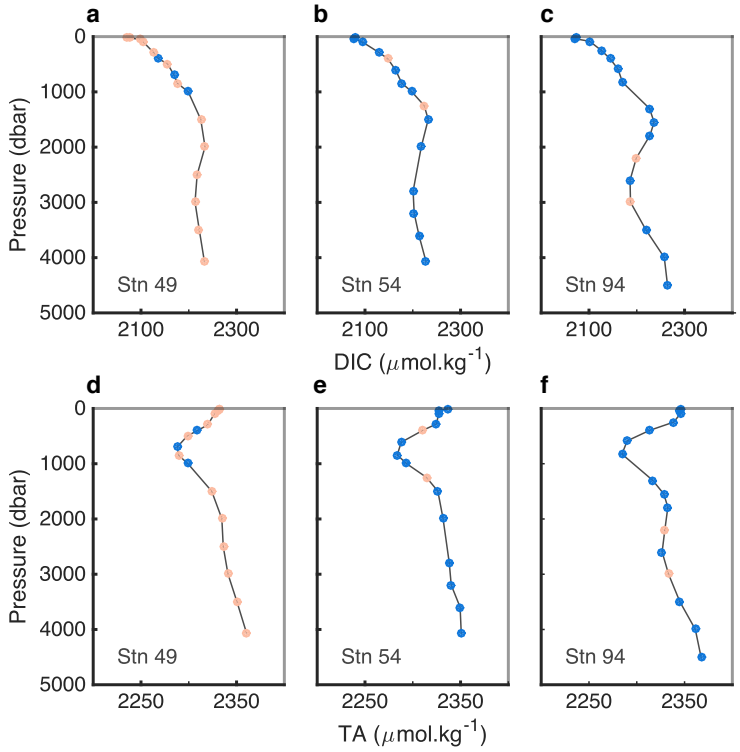

Figure B.4: Testing the calibration adjustments between VINDTA #064 and #065. Vertical profiles of **a-c**, DIC, **d-f** TA for three different stations: 49 (**a,d**), 54 (**b,e**) and 94 (**c,f**). Those stations were chosen as their profiles were measured on both machines: VINDTA #064 (blue) and #065 (pink).

B.5 Second quality control

The second level QC aims to identify potential systematic biases through a crossover analysis, which broadly compares profiles at depth (below 1,500 m) and within a 2° arc-distance with historical measurements (Lauvset and Tanhua, 2015). The MSM060 crossover analysis identified 11 past voyages, 5 of which were subjected to DIC and sometimes TA adjustments (Table B.2). For instance, the voyage 316N19871123.1 overestimated the reference DIC background by 19 µmol·kg^{-1} and was therefore corrected by the GLODAP community by subtracting 19 µmol·kg^{-1} from all DIC data collected

during that voyage. Note that most of the adjustments, as well as the highest adjustment amplitudes, were applied to voyages from the pre-CRM era (i.e. prior to 1995).

Label	EXPO code	Year	TA / DIC adjustments (µmol·kg^{-1})	Comments
(A)	316N19831007	1983	+4 / +8	-
(B)	316N19871123.1	1989	-19 / -11	-
(C)	323019940104	1994	-12 / 0	-
(D)	35A319950113	1995	-6 / +3	-
(E)	35A319950221	1995	0 / 0	-
(F)	90AV20041104	2004	0 / -	-
(G)	33RO20050111	2005	+5 / 0	-
(H)	35MF20080207	2008	0 / 0	DIC might be slightly low ($\sim$2)
(I)	33RO20100308	2010	0 / 0	-
(J)	29HE20100208	2010	0 / -	-
(K)	29HE20130320	2013	0 / -	TA biased low

Table B.2: List of voyages used in the crossover analysis. The DIC and TA adjustments are taken from the OCADS online adjustment table (https://glodapv2.geomar.de/adjustments/list). Note that in some cases, the adjustment table was unclear which final adjustment values were used in GLODAPv2; the values indicated here are therefore the ones we *think* were applied.

In a DIC crossover example (Figure B.5 left), 13 stations and 68 samples from MSM060 are crossed with 14 stations and 119 samples from voyage 35MF20080207 along the coast of South Africa. Since the weighted mean offset (i.e. -2.14 µmol·kg^{-1}) is within the limit levels of $\pm$ 4 µmol·kg^{-1}, the crossover analysis with voyage 35MF20080207 suggests that our data do not contain any systematic bias. However, this conclusion could be altered by the high standard deviation value (i.e. 8.62 µmol·kg^{-1}, which seems to be highly influenced by the offset of station 12 (Figure B.5 left). To study the impact of this station on the crossover analysis' results, the above study is repeated once station 12 removed (Figure B.5 right) and concludes that the offset in the weighted mean and the standard deviation is reduced (i.e. -0.63 $\pm$ 6.3 µmol·kg^{-1}). While no obvious instrumental issues occurred during the DIC measurements of station 12, the suggested offset may not be due to an accuracy issues, but rather the presence of a physical feature such as an eddy.

Using the same approach as in this example, four "suspicious" profiles were identified: the DIC profiles for stations 12 and 104, and the TA profiles for stations 15 and 104. Note that those four profiles, which are in the western and eastern sides of the basin, were identified because they were crossed over with previous voyages. The central part of the basin might as well have some relatively odd profiles, but just happened to not be included in the crossover

analysis and therefore not identified here.

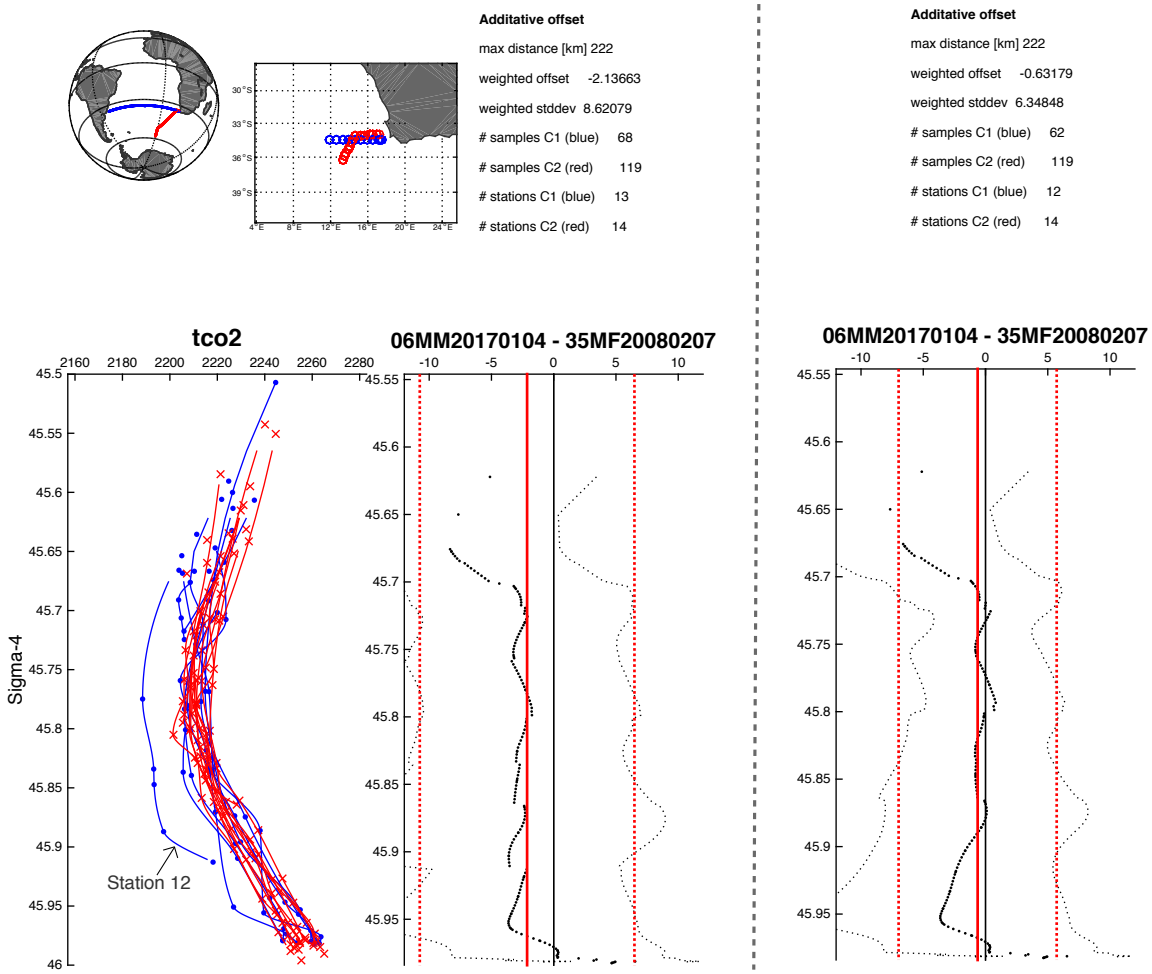

Figure B.5: Example of a DIC crossover analysis. Results of a crossover analysis between the MSM060 voyage (blue) and the reference voyage 35MF20080207 (red). The top part includes the complete track of those two voyages, the zoomed map with the selected stations for the crossover analysis, and the main information about this crossover (weighted mean offset, weighted standard deviation, number of samples and stations used from both voyages). On the left of the grey dashed line is: the DIC profiles along the density sigma levels for the two voyages, with the profile of station 12 indicated (left) and the offsets in the weighted mean and standard deviation profiles between the two voyages (right). On the right of the grey dashed line is the equivalent but after removing the suspicious profile of station 12.

The overall crossover results (Table B.3; Figure B.6) shows that the voyage MSM060 does not present any systematic bias in both the DIC (1.87 ± 2.04 µmol·kg^{-1}) and TA measurements (1.85 ± 0.95 µmol·kg^{-1}). However, crossovers results using voyages prior the CRM era present a higher weighted mean offset compared to the post CRM era voyages (Table B.3), suggesting that potential systematic biases main remain in those early voyages

and that further adjustments by the GLODAP community could be made. Out of curiosity, the impact of the four relatively odd profiles identified above on the overall crossover results was investigated. After removing the four suspicious profiles, the weighted mean offset is reduced for the crossover analysis on TA but not for the DIC analysis. However all results (i.e. pre-CRM, post-CRM and all; DIC and TA) show a reduction in the standard deviation from the offset weighted mean (Table B.3).

Despite the apprehension related to the challenges found with the CRM calibration (i.e. the high offset returned by the VINDTAs), the second level QC confirms that the DIC and TA data produced during the MSM060 voyage reply to the GO-SHIP high-quality standards.

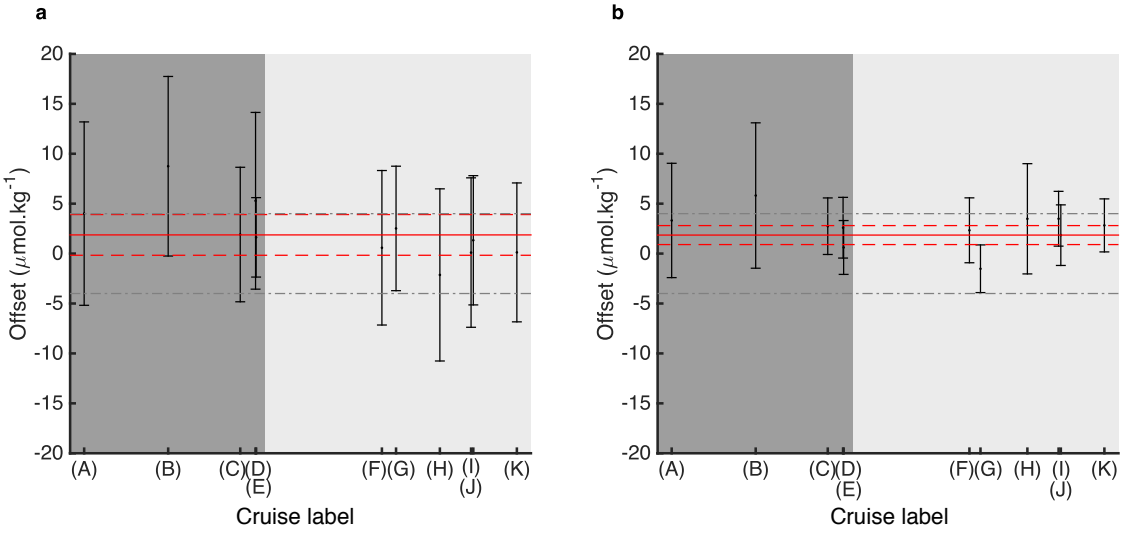

Figure B.6: Individual voyage offset returned by the crossover analysis. Weighted mean of the absolute offset and standard deviation of weighted mean returned by the **a**, DIC and **b**, TA crossover analyses, separated using the voyages from the pre-CRM era (dark grey in Figure B.6), post-CRM era (light grey in Figure B.6), and all voyages (red lines in Figure B.6). A positive offset means that the voyage MSM060 might have a positive bias compared to the reference voyage. Voyage label is referred in Table B.2.

	Pre-CRM era	Post-CRM era	All
DIC (μmol·kg^{-1})	3.01 ± 2.86	0.69 ± 2.91	1.87 ± 2.04
	(3.98 ± 2.57)	(1.41 ± 2.40)	(2.61 ± 1.75)
TA (μmol·kg^{-1})	2.19 ± 1.54	1.65 ± 1.21	1.85 ± 0.95
	(2.08 ± 1.51)	(1.32 ± 1.15)	(1.60 ± 0.92)

Table B.3: Crossover analysis summary. Weighted mean of the absolute offset and standard deviation of weighted mean returned by the DIC and TA crossover analyses, separated using the voyages from the pre-CRM era (dark grey in Figure B.6), post-CRM era (light grey in Figure B.6), and all voyages (red lines in Figure B.6). Results from the crossover analysis in which the 4 suspicious profiles were removed (stations 12 and 104 for DIC and stations 15 and 104 for TA) are shown in brackets.

Acknowledgments

This is surely the most difficult part for me to write. Not because I don't know who to thank, but rather I feel that the words will never fully capture how grateful I am. Becoming a climate scientist has been a dream of mine since I was 14. This Ph.D. is therefore not just about the people who've helped me during those 3.5 years, but mostly about the people that have been there along the way. And of course, I think of my family. My mum especially. She is my lifetime example of strength and bravery. Science is miles away from her world, but yet her support for me to pursue my goal has been generous and unconditional. I know how proud of me she is, but I hope that finally she's going to be proud of herself, not only because she is amazing, but because despite all the things life threw at her, at us, she did it. (Maman, merci pour tous tes sacrifices, ta confiance et ta force surtout, une source d'inspiration au quotidien. Je t'en serai à vie reconnaissante.) My sister and aunties have also been there from the beginning. I am grateful for their unconditional support, their patient ears and helpful advice. This one is for you too, grandma.

I also want to thank my people, the ones that have been around when my ship is rocking in stormy seas or simply cruising under the sun. And here they are: Estelle, Clara, Claire, Maria, Yuna, Bertille, Matt, Steve, Tobia, Sarah, Karmen, Rama, Lailah, Shakara, Oumou, Lucy, Leticia, Antou, Mary, Precious, Luke, Amanda. All there in different ways, at different times, but all there. To some we shared a flat, to others a lifestyle, a passion, a work environment or a long friendship, but all showed me kindness and I am grateful for that. Of course, I want to thank my other family, my Street Family. They've all been a major source of energy and allowed my mind to escape from time to time, and come back more focused, more powerful. And at the source, you, Patrice, who've also been on this journey with me for a while. My spiritual support. Unconsciously surely, but there always.

I am now coming to my dear colleagues and supervisors. First, I would like to thank the

members of the jury, Laurent Bopp and Pierre Friedlingstein, for taking some of their precious time for my work; I am honoured and hope you enjoyed the read (if you've made it that far, well done!). I would like then to thank David Ford, Siv Lauvset, Peter Landschützer, Peter Cox and Pedro Monteiro for their help, great contributions and encouragement to my first (unfinished) publication journey - what a ride. Finally, I am very much grateful to my supervisors, Andy, Ute and Doug. I feel lucky to have learnt from each of you. And of course, from you, Paul too. It was a real honour to be your first Ph.D. student, and I hope the feeling is reciprocal. I deeply thank you for taking me under your wing, nurturing me with wise advice, patience, honesty and with kindness always. You saw me grow, and hopefully now you can see me fly. And maybe the one above would be proud to see me rise.

A.D.L.

Bibliography

Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Shindo, E., Tsujino, H., Mizuta, R., S., Y., Koshiro, T., Ose, T., and Kitoh, A. (2013). Basic performance of a new earth system model of the Meteorological Research Institute (MRI-ESM1). *Papers in Meteorology and Geophysics*, 64:1–19.

Amante, C. and Eakins, B. (2015). ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. Technical report, National Oceanic Atmospheric Administration.

Anderson, T. R. (2005). Plankton Functional Type modelling: running before we can walk? *Journal of Plankton Research*, 27(11):1073–1081.

Arruda, R., Calil, P. H., Bianchi, A. A., Doney, S. C., Gruber, N., Lima, I., and Turi, G. (2015). Air-sea CO_2 fluxes and the controls on ocean surface pCO_2 seasonal variability in the coastal and open-ocean southwestern Atlantic Ocean: A modeling study. *Biogeosciences*, 12(19):5793–5809.

Aumont, O. and Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, 20(2):1–15.

Bakker, D. C., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S. I., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W. J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S.,

Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., Van Heuven, S. M., Vandemark, D., Ward, B., Watson, A. J., and Xu, S. (2016). A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT). *Earth System Science Data*, 8(2):383–413.

Bakker, D. C. E., Pfeil, B., Smith, K., Hankin, S., Olsen, A., Alin, S. R., Cosca, C., Harasawa, S., Kozyr, A., Nojiri, Y., O'Brien, K. M., Schuster, U., Telszewski, M., Tilbrook, B., Wada, C., Akl, J., Barbero, L., Bates, N. R., Boutin, J., Bozec, Y., Cai, W. J., Castle, R. D., Chavez, F. P., Chen, L., Chierici, M., Currie, K., De Baar, H. J. W., Evans, W., Feely, R. A., Fransson, A., Gao, Z., Hales, B., Hardman-Mountford, N. J., Hoppema, M., Huang, W. J., Hunt, C. W., Huss, B., Ichikawa, T., Johannessen, T., Jones, E. M., Jones, S. D., Jutterström, S., Kitidis, V., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Manke, A. B., Mathis, J. T., Merlivat, L., Metzl, N., Murata, A., Newberger, T., Omar, A. M., Ono, T., Park, G. H., Paterson, K., Pierrot, D., Ríos, A. F., Sabine, C. L., Saito, S., Salisbury, J., S. Sarma, V. V. S., Schlitzer, R., Sieger, R., Skjelvan, I., Steinhoff, T., Sullivan, K. F., Sun, H., Sutton, A. J., Suzuki, T., Sweeney, C., Takahashi, T., Tjiputra, J., Tsurushima, N., C. Van Heuven, S. M. a., Vandemark, D., Vlahos, P., Wallace, D. W. R., Wanninkhof, R., and Watson, A. J. (2014). An update to the surface ocean CO_2 atlas (SOCAT version 2). *Earth System Science Data*, 6:69–90.

Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P., and White, J. W. C. (2012). Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature*, 488(7409):70–72.

Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R., Anderegg, W., Cooper, L. A., DeGrandpre, M., Tans, P. P., Miller, J. B., Alden, C., and White, J. W. (2015). Audit of the global carbon budget: Estimate errors and their impact on uptake uncertainty. *Biogeosciences*, 12(8):2565–2584.

Bates, N. R., Best, M. H. P., Neely, K., Garley, R., Dickson, A. G., and Johnson, R. J. (2012). Detecting anthropogenic carbon dioxide uptake and ocean acidification in the North Atlantic Ocean. *Biogeosciences*, 9:2509–2522.

Bates, R. N., Astor, Y. M., J., C. M., Currie, K., Dore, J. E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafs-

son, J., and Santana-Casiano, J. M. (2014). A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic CO_2 and ocean acidification. *Oceanography*, 27(1):126–141.

Beckers, J. M., Barth, A., Troupin, A., and Alvera-Azcárate, A. (2014). Approximate and efficient methods to assess error fields in spatial gridding with data interpolating variational analysis (DIVA). *Journal of Atmospheric and Oceanic Technology*, 31(2):515–530.

Bennington, V., McKinley, G. A., Dutkiewicz, S., and Ullman, D. (2009). What does chlorophyll variability tell us about export and air-sea CO_2 flux variability in the North Atlantic? *Global Biogeochemical Cycles*, 23(3):1–11.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevå g, A., Seland, O., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E. (2013). The Norwegian Earth System Model, NorESM1-M, Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, 6(3):687–720.

Boden, T., Marland, G., and Andres, R. (2013). Global, regional, and national fossil-fuel CO_2 emissions. Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N. (2012). Aerosols implicated as a prime driver of twentieth-century north atlantic climate variability. *Nature*.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M. (2013). Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*, 10(10):6225–6245.

Buesseler, K. O. (1998). The decoupling of production and particulate export in the surface ocean. *Global Biogeochemical Cycles*, 12:297–310.

Canadell, J. G., Le Quere, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., and Marland, G. (2007). Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences*, 104(47):18866–18870.

Chylek, P., Li, J., Dubey, M. K., Wang, M., and Lesins, G. (2011). Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2. *Atmospheric Chemistry and Physics Discussions*, 11(8):22893–22907.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R., Piao, S., and Thornton, P. (2013). Carbon and other biogeochemical cycles. in: Climate change 2013: The physical science basis. Technical report, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S. (2011). Development and evaluation of an Earth-System model – HadGEM2. *Geoscientific Model Development*, 4(4):1051–1075.

Cooper, D. J., Watson, A. J., and Ling, R. D. (1998). Variation of p(CO2) along a North Atlantic shipping route (U.K. to the Caribbean): A year of automated observations. *Marine Chemistry*, 60(1-2):147–164.

Couldrey, M. P., Oliver, K. I., Yool, A., Halloran, P. R., and Achterberg, E. P. (2016). On which timescales do gas transfer velocities control North Atlantic CO_2 flux variability? *Global Biogeochemical Cycles*, 30(5):787–802.

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., and Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, 494(7437):341–344.

Craig, H. and Gordon, A. (1972). The GEOSECS program: 1970-1971. *Earth and planetary science letters*, 16:47–49.

Cubasch, U., Wuebbles, D., Chen, D., Facchini, M., Frame, D., Mahowald, N., and Winther, J.-G. (2013). Introduction. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Technical report, Cambridge

University Press, Cambridge, United Kingdom and New York, NY, USA.

Cunningham, S. A., Roberts, C. D., Frajka-Williams, E., Johns, W. E., Hobbs, W., Palmer, M. D., Rayner, D., Smeed, D. A., and McCarthy, G. (2013). Atlantic Meridional Overturning Circulation slowdown cooled the subtropical ocean. *Geophysical Research Letters*, 40(23):6202–6207.

De La Rocha, C. L. (2007). *Biogeochemistry: Treatise on geochemistry. The biological pump*, volume 8.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorological Society*, 137:553–597.

Denman, K., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P., Dickinson, R., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P., Wofsy, S., and Zhang, X. (2007). *Couplings Between Changes in the Climate System and Biogeochemistry*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

DeVries, T., Holzer, M., and Primeau, F. (2017). Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning. *Nature*, 542(7640):215–218.

Dickson, A. (1981). An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data. *Deep Sea Research*, 28A:609–623.

Dickson, A. G., Sabine, C. L., and Christian, J. R. (2007). *Guide to best practices for ocean CO_2 measurements*. PICES Special Publication 3, IOCCP REPORT.

Dlugokencky, E. and Tans, P. (2016). NOAA/ESRL.

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead,

L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N. (2013). *Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5*, volume 40.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A., and Zadeh, N. (2013). GFDLs ESM2 global coupled climate-carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics. *Journal of Climate*, 26:2247–2267.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A., and Zadeh, N. (2012). GFDLs ESM2 global coupled climate-carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation characteristics. *Journal of Climate*, 21:6646–6665.

Etheridge, D. (2010). Law Dome Ice Core 2000-Year CO_2, CH_4, and N_2O Data.

Etheridge, D., Steele, L., Francey, R., and Langenfelds, R. (1998). Atmospheric methane between 1000 A.D. and present: evidence of anthropogenic emissions and climatic variability. *Journal of Geophysical Research*, 103:15979–15996.

Etheridge, D., Steele, L., Langenfelds, R., Francey, R., Barnola, J.-M., and Morgan, V. (1996). Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn. *Journal of Geophysical Research*, 101:4115–4128.

Fay, A. R. and McKinley, G. A. (2013). Global trends in surface ocean pCO_2 from in situ data. *Global Biogeochemical Cycles*, 27(2):541–557.

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J. (2004). Impact of Anthropogenic CO_2 on the $CaCO_3$ System in the Oceans. *Science*, 305(5682):362–366.

Ford, D. and Barciela, R. (2017). Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products. *Remote Sensing of Environment*, (2016).

Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., and Demaria, J. (2012). Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Science*, 8(5):751–771.

Friedrichs, M. A., Carr, M. E., Barber, R. T., Scardi, M., Antoine, D., Armstrong, R. A., Asanuma, I., Behrenfeld, M. J., Buitenhuis, E. T., Chai, F., Christian, J. R., Ciotti, A. M., Doney, S. C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J., Kameda, T., Lima, I., Marra, J., Mélin, F., Moore, J. K., Morel, A., O'Malley, R. T., O'Reilly, J., Saba, V. S., Schmeltz, M., Smyth, T. J., Tjiputra, J., Waters, K., Westberry, T. K., and Winguth, A. (2009). Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *Journal of Marine Systems*, 76(1-2):113–133.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., and Antonov, J. I. (2010). World Ocean Atlas 2009, vol. 4, Nutrients (Phosphate, Nitrate, Silicate).

Geider, R. J., MacIntyre, H. L., and Kana, T. M. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *Limnology and Oceanogry*, 43(4):679–694.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z. L., and Zhang, M. (2011). The Community Climate System Model version 4. *Journal of Climate*, 24(19):4947–4991.

Giorgetta, M. A., Jungclaus, J. H., Reick, C. H., Legutke, S., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H. D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W. A., Notz, D., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K., Stockhause, M., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., , and Stevens, B. (2013). Climate

change from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project 5. *Journal of Advances in modeling earth systems*, 5(3):572–597.

GLOBALVIEW-CO2 (2013). Multi-laboratory compilation of synchronized and gap-filled atmospheric carbon dioxide records for the period 1979-2012. Technical report, NOAA Global Monitoring Division: Boulder, Colorado, U.S.A.

Good, S. A., Martin, M. J., and Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118:6704–671.

Gouretski, V. and Reseghetti, F. (2010). On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep Sea Research Part I: Oceanographic Research Papers*, 57(6):812–833.

Gregory, J. M., Dixon, K. W., Stouffer, R. J., Weaver, A. J., Driesschaert, E., Eby, M., Fichefet, T., Hasumi, H., Hu, A., Jungclaus, J. H., Kamenkovich, I. V., Levermann, A., Montoya, M., Murakami, S., Nawrath, S., Oka, A., Sokolov, A. P., , and Thorpe, R. B. (2005). A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 concentration. *Geophysical Research Letters*, 23(32):L12703.

Gruber, N., Gloor, M., Mikaloff, F., E., S., Doney, S. C., Dutkiewicz, S., Follows, M. J., Gerber, M., Jacobson, A. R., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S., Sarmiento, J. L., and Takahashi, T. (2009). Oceanic sources, sinks, and transport of atmospheric CO_2. *Global Biogeochemical Cycles*, 23(1):1–21.

Halloran, P. R., Bell, T. G., and Totterdell, I. J. (2010). Can we trust empirical marine DMS parameterisations within projections of future climate? *Biogeosciences*, 7:1645–1656.

Halloran, P. R., Booth, B. B. B., Jones, C. D., Lambert, F. H., McNeall, D. J., Totterdell, I. J., and Volker, C. (2015). The mechanisms of North Atlantic CO_2 uptake in a large Earth System Model ensemble. *Biogeosciences*, 12:4497–4508.

Hansen, J., Sato, M., Ruedy, R., Kharecha, P., Lacis, A., Miller, R., Nazarenko, L., Lo, K., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, S., Baum, E., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Cohen, A., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Jonas, J.,

Kelley, M., Kiang, N. Y., Koch, D., Labow, G., Lerner, J., Menon, S., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Schmunk, R., Shindell, D., Stone, P., Sun, S., Streets, D., Tausnev, N., Thresher, D., Unger, N., Yao, M., and Zhang, S. (2007). Climate simulations for 1880-2003 with GISS model. *Climate Dynamics*, 29(7-8):661–696.

Hastie, T., Tibshirani, R., and Friedman, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd Edition.

Houghton, R., van der Werf, G., DeFries, R., Hansen, M., J.I., H., Le Quéré, C., Pongratz, J., and Ramankutty, N. (2012). Carbon emissions from land use and land-cover change. *Biogeosciences*, 9(12):5125–5142.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., and Elliott, S. (2010). CICE: the Los Alamos Sea Ice Model Documentation and Software User's Manual Version 4.1.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M. (2003). An overview of the North Atlantic Oscillation. *Geophysical Monograph*, 134:1–34.

Iida, Y., Kojima, A., Takatani, Y., Nakano, T., Sugimoto, H., Midorikawa, T., and Ishii, M. (2015). Trends in pCO_2 and sea-air CO_2 flux over the global open oceans for the last two decades. *Journal of Oceanography*, 71(6):637–661.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., and Nunez-Riboni, I. (2013). The global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth System Model in different CMIP5 experimental realizations. *Geophysical Research Letters*, 40:5909–5914.

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A. M., Helber, R., and Arnone, R. A. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, 76(1-2):64–82.

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., Brovkin, V., Hajima, T., Kato, E., Kawamiya, M., Liddicoat, S., Lindsay, K., Reick, C. H., Roelandt, C., Segschneider, J., and Tjiputra, J. (2013). Twenty-First-Century Compatible CO_2 Emissions and Airborne Fraction Simulated by CMIP5 Earth System Models under Four Representative Concentration Pathways. *Journal of Climate*, 26(13):4398–4413.

Jones, S. D., Le Quéré, C., Rödenbeck, C., Manning, A. C., and Olsen, A. (2015). A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data. *Journal of Advances in Modeling Earth Systems*, 7:1554–1575,.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S. (2013). Characteristics of the ocean simulations in the Max Planck InstituteOcean Model (MPIOM) the ocean component of the MPI Earth System Model. *Journal of Advances in Modeling Earth Systems*, 5(21):422–446.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–471.

Kara, A. B., Rochford, P. A., and Hurlburt, H. E. (2000). An optimal definition for ocean mixed layer depth. *Journal of Geophysical Research*, 105(C7):16803.

Kawamiya, M., Kishi, M. J., and Suginohara, N. (2000). An ecosystem model for the North Pacific embedded in a general circulation model Part II: Mechanisms forming seasonal variations of chlorophyll. *Journal of Marine Systems*, 25:159–178.

Keenan, T. F., Prentice, I. C., Canadell, J. G., Williams, C. A., Wang, H., Raupach, M., and Collatz, G. J. (2016). Recent pause in the growth rate of atmospheric CO_2 due to enhanced terrestrial carbon uptake. *Nature Communications*, 7(13428).

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T. H. (2004). A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles*, 18(4):1–23.

Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., and Suzuki, T. (2015). Global Ocean Data Analysis Project, Version 2 (GLODAPv2). *ORNL/CDIAC-162*, NDP-093.

Khatiwala, S., Tanhua, T., Mikaloff Fletcher, S., Gerber, M., Doney, S. C., Graven, H. D., Gruber, N., McKinley, G. A., Murata, A., Ríos, A. F., and Sabine,

C. L. (2013). Global ocean storage of anthropogenic carbon. *Biogeosciences*, 10(4):2169–2191.

Kortzinger, A. (1999). *Determination of carbon dioxide partial pressure (p(CO2)), in Methods of Seawater Analysis*.

Kortzinger, A., Send, U., Lampitt, R. S., Hartman, S., Wallace, D. W. R., Karstensen, J., Villagarcia, M. G., Llinas, O., and DeGrandpre, M. D. (2008). The seasonal pCO_2 cycle at 49○N/16.5○W in the northeastern Atlantic Ocean and what it tells us about biological productivity. *Journal of Geophysical Research*, 113(4):1–15.

Kuhlbrodt, T., Griesel, a., Montoya, M., Levermann, a., Hofmann, M., and Rahmstorf, S. (2007). On the driving processes of the Atlantic meridional overturning circulation. *Atlantic*, 45(2004):RG2001.

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Li, Y., and Séférian, R. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change*, 7:355–358.

Landschützer, P., Gruber, N., and Bakker, D. C. E. (2015). A 30 years observation-based global monthly gridded sea surface pCO_2 product from 1982 through 2011.

Landschützer, P., Gruber, N., and Bakker, D. C. E. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, 30(10):1396–1417.

Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U. (2014). Recent variability of the global ocean carbon sink. *Global Biogeochemical Cycles*, 28:927–949.

Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, 10(11):7793–7815.

Lauvset, S. K., Key, R. M., Olsen, A., Van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S. (2016). A new global interior ocean mapped climatology: The $1^\circ \times 1^\circ$ GLODAP version 2. *Earth System Science Data*, 8(2):325–340.

Lauvset, S. K. and Tanhua, T. (2015). A

toolbox for secondary quality control on ocean chemistry and hydrographic data. *Limnology and Oceanography: Methods*, 13(11):601–608.

Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I., Keeling, R. F., Tans, P., Arneth, A., Bakker, D. C., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R., Murata, A., S. Nabel, J. E., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., Pérez, F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Séférian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., Van Der Laan-Luijkx, I. T., Van Der Werf, G. R., Van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S., and Zeng, N. (2015). Global Carbon Budget 2015. *Earth System Science Data*, 7(2):349–396.

Le Treut, H., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., and Prather, M. (2007). Historical overview of climate change. in: Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change [solomon, s., d. qin, m. manning, z. chen, m. marquis, k.b. averyt, m. tignor and h.l. miller (eds.)]. Technical report, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Lefèvre, N., Watson, A. J., Olsen, A., Ríos, A. F., Pérez, F. F., and Johannessen, T. (2004). A decrease in the sink for atmospheric CO_2 in the North Atlantic. *Geophysical Research Letters*, 31(7):2–5.

Liss, P. S. and Merlivat, L. (1986). *Air-sea gas exchange rates: Introduction and synthesis, in: The role of air-sea exchange in geochemical cycling*. Springer.

Long, M., Lindsay, K., and Peacock, S. (2011). Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1(BGC)*. *Journal of Climate*, 26(3):6775–6800.

Longhurst, A. R. (2007). *Ecological Geography of the Sea (2nd Edition)*. Academic Press, London.

Lueker, T., Dickson, A., and Keeling, C.

(2000). pCO_2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: validation based on laboratory measurements of CO_2 in gas and seawater at equilibrium. *Marine Chemistry*, pages 105–119.

MacFarling Meure, C., Etheridge, D., Trudinger, C., Steele, P., Langenfelds, R., van Ommen, T., Smith, A., and Elkins, J. (2006). The Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP. *Geophysical Research Letters*, 33(14):L14810.

Madec, G. (2008). *NEMO ocean engine*. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619.

Marotzke, J. and Forster, P. M. (2015). Forcing, feedback and internal variability in global temperature trends. *Nature*, 517(7536):565–570.

Marshall, J. and Schott, F. (1999). Open-ocean convection: observations, theory and models. *Reviews of Geophysics*, 37:1–64.

McDougall, T. J. and Barker, P. M. (2011). Getting started with TEOS 10 and the Gibbs Seawater (GSW) Oceanographic Toolbox. 28pp, SCOR/IAPSO WG127, ISBN 978-0-646-55621-5.

McKinley, G. A., Fay, A. R., Takahashi, T., and Metzl, N. (2011). Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales. *Nature Geoscience*, 4(9):606–610.

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., and Lovenduski, N. S. (2016). Timescales for detection of trends in the ocean carbon sink. *Nature*, 530(7591):469–472.

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., Hyder, P., Siddorn, J., and Sinha, B. (2014). GO5. 0: The joint NERC-Met Office NEMO global ocean model for use in coupled and forced applications. *Geoscientific Model Development*, 7(3):1069–1092.

Menemenlis, D., Campin, J., Heimbach, P., Hill, C., Lee, T., Nguyen, A., Schodlok, M., and Zhang, H. (2008). ECCO2: High resolution global ocean and sea ice data synthesis. *Mercator Ocean Quarterly Newsletter*, 31(October):13–21.

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Muller, S. A., and Sarmiento, J. L. (2007). Inverse estimates of the oceanic sources and sinks of natural CO_2

and the implied oceanic carbon transport. *Global Biogeochemical Cycles*, 21(1):1–19.

Millero, F., Pierrot, D., Lee, K., Wanninkholf, R., Feely, R., Sabine, C., Key, R., and Takahashi, T. (2002). Dissociation constants for carbonic acid determined from field measurements. *Deep Sea Research I*, 49:1705–1723.

Mongwe, N. P., Chang, N., and Monteiro, P. (2016). The seasonal cycle as a mode to diagnose biases in modelled CO_2 fluxes in the Southern Ocean. *Ocean Modelling*, 106:90–103.

Monteiro, P. M. S., Gregor, L., Lëvy, M., Maenner, S., Sabine, C. L., and Swart, S. (2015). Intraseasonal variability linked to sampling alias in air-sea CO_2 fluxes in the Southern Ocean. *Geophysical Research Letter*, 42.

Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J., and Upsill-Goddard, R. C. (2000). In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers. *Global Biogeochemical Cycles*, 14(1):373–387.

Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F. F., and Suzuki, T. (2016). The global ocean data analysis project version 2 (GLODAPv2) - An internally consistent data product for the world ocean. *Earth System Science Data*, 8(2):297–323.

Palmer, J. R. and Totterdell, I. J. (2001). Production and export in a global ecosystem model. *Deep-Sea Research Part I: Oceanographic Research Papers*, 48(5):1169–1198.

Pérez, F. F., Mercier, H., Vázquez-Rodríguez, M., Lherminier, P., Velo, A., Pardo, P. C., Rosón, G., and Ríos, A. F. (2013). Atlantic Ocean CO_2 uptake reduced by weakening of the meridional overturning circulation. *Nature Geoscience*, 6(2):146–152.

Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, J., Manke, A., Metzl, N., Sabine, C. L., Akl, J., Alin, S. R., Bates, N., Bellerby, R. G. J., Borges, A., Boutin, J., Brown, P. J., Cai, W. J., Chavez, F. P., Chen, A., Cosca, C., Fassbender, A. J., Feely, R. A., González-Dávila, M., Goyet, C., Hales, B., Hardman-Mountford, N., Heinze, C., Hood, M., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannessen, T., Jones,

S. D., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Nakadate, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G. H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A., Ríos, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V. S. S., Schlitzer, R., Schneider, B., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Tjiputra, J., Vandemark, D., Veness, T., Wanninkhof, R., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H. (2013). A uniform, quality controlled Surface Ocean CO_2 Atlas (SOCAT). *Earth System Science Data*, 5(1):125–143.

Pickart, R. S., Spall, M. A., Ribergaard, M. H., Moore, G. W. K., and Milliff, R. F. (2003). Deep convection in the Irminger Sea forced by the Greenland tip jet. *Nature*, 424:152–156.

Pierrot, D., Neill, C., Sullivan, K., Castle, R., Wanninkhof, R., Luüger, H., Johannessen, T., Olsen, A., Feely, R. A., and Cosca, C. E. (2009). Recommendations for autonomous underway pCO_2 measuring systems and data-reduction routines. *Deep-Sea Research II*, 56:512–522.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D., Casey, K. S., and Schlax, M. G. (2007). Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate*, 20(22):5473–5496.

Riahi, K., Grübler, A., and Nakicenovic, N. (2007). Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change*, 74(7):887–935.

Robertson, J. E. and Watson, A. J. (1992). Thermal skin effect of the surface ocean and its implications for CO_2 uptake. *Nature*, 358:738–740.

Romanou, A., Romanski, J., and Gregg, W. W. (2014). Natural ocean carbon cycle sensitivity to parameterizations of the recycling in a climate model. *Biogeosciences*, 11:1137–1154.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, W. R., Tilbrook, B., Millero, F. J., Peng, T., Kozyr, A., Ono, T., and Rios, A. F. (2004).

The Oceanic Sink for Anthropogenic CO_2. *Science*, 305(5682):367–371.

Sabine, C. L., Hankin, S., Koyuk, H., Bakker, D. C., Pfeil, B., Olsen, A., Metzl, N., Kozyr, A., Fassbender, A., Manke, A., Malczyk, J., Akl, J., Alin, S. R., Bellerby, R. G., Borges, A., Boutin, J., Brown, P. J., Cai, W. J., Chavez, F. P., Chen, A., Cosca, C., Feely, R. A., González-Dávila, M., Goyet, C., Hardman-Mountford, N., Heinze, C., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannessen, T., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Nakadate, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G. H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A., Ríos, A. F., Salisbury, J., Santana-Casiano, J. M., S. Sarma, V. V., Schlitzer, R., Schneider, B., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Vandemark, D., Veness, T., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H. (2013). Surface Ocean CO_2 Atlas (SOCAT) gridded data products. *Earth System Science Data*, 5(1):145–153.

Sarmiento, J. L. and Gruber, N. (2006). *Ocean Biogeochemical Dynamics*. Princeton University Press.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., Genio, A. D., de Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M., and Zhang, J. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, 6:141–184.

Schuster, U., McKinley, G. A., Bates, N., Chevallier, F., Doney, S. C., Fay, A. R., González-Dávila, M., Gruber, N., Jones, S., Krijnen, J., Landschützer, P., Lefèvre, N., Manizza, M., Mathis, J., Metzl, N., Olsen, A., Rios, A. F., Rödenbeck, C., Santana-Casiano, J. M., Takahashi, T., Wanninkhof, R., and Watson, A. J. (2013). An assessment of the Atlantic and Arctic sea–air CO_2 fluxes, 1990-2009. *Biogeosciences*, 10(1):607–627.

Schuster, U. and Watson, A. J. (2007). A variable and decreasing sink for atmospheric CO_2 in the North Atlantic. *Journal of Geophysical Research*, 112(11):C11006.

Schuster, U., Watson, A. J., Bakker, D. C. E., De Boer, A. M., Jones, E. M., Lee, G. A., Legge, O., Louwerse, A., Riley, J., and Scally, S. (2014). Measurements of total alkalinity and inorganic dissolved carbon in the Atlantic Ocean and adjacent Southern Ocean between 2008 and 2010. *Earth System Science Data*, 6(1):175–183.

Schuster, U., Watson, A. J., Bates, N. R., Corbiere, A., Gonzalez-Davila, M., Metzl, N., Pierrot, D., and Santana-Casiano, M. (2009). Trends in North Atlantic sea-surface fCO_2 from 1990 to 2006. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 56(8-10):620–629.

Schutler, J. D., Land, P. E., Piolle, J., Woolf, D. K., Goddijn-Murphy, L., Paul, F., Girard-Ardhuin, F., Chapron, B., and Donlon, C. J. (2016). FluxEngine: A Flexible Processing System for Calculating Atmosphere-Ocean Carbon Dioxide Gas Fluxes and Climatologies. *Journal of Atmospheric and Oceanic Technology*, 33(1):741–756.

Scott, V., Kettle, H., and Merchant, C. J. (2011). Sensitivity analysis of an ocean carbon cycle model in the North Atlantic: An investigation of parameters affecting the air-sea CO_2 flux, primary production and export of detritus. *Ocean Science*, 7(3):405–419.

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., Dunne, J. P., Christian, J. R., Doney, S. C., Ilyina, T., Lindsay, K., Halloran, P. R., Heinze, C., Segschneider, J., Tjiputra, J., Aumont, O., and Romanou, A. (2016). Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, 9(5):1827–1851.

Sims, R. P., Schuster, U., Watson, A. J., Yang, M. X., Hopkins, F. E., Stephens, J., and Bell, T. G. (2017). A measurement system for vertical seawater profiles close to the air-sea interface. *Ocean Science*, 13(5):649–660.

Stramma, L., Kiebe, D., Rhein, M., Schott, F., Yashayaev, I., and Koltermann, K. P. (2004). Deep water changes at the western boundary of the subpolar North Atlantic during 1996 to 2001. *Deep-Sea Research, Part I*, (51):1033–1056.

Stuiver, M., Quay, P. D., and Ostlund, H. G. (1983). Abyssal water carbon-14 distribu-

tion and the age of the world oceans. *Science*, 219(4586):849–851.

Sudo, K., Takahashi, M., Kurokawa, J., , and Akimoto, H. (2002). CHASER: A global chemical model of the troposphere 1. Model description. *Journal of Geophysical Research*, 107:4339.

Sun, S. and Bleck, R. (2006). Multi-century simulations with the coupled GISS-HYCOM climate model: Control experiments. *Climate Dynamics*, 26(4):407–428.

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W., and Sutherland, S. (1993). Seasonal variation of CO_2 and nutrients in the high-latitude surface oceans: A comparative study. *Global Biogeochemical Cycles*, 7(4):843–878.

Takahashi, T. and Sutherland, S. C. (2013). Global ocean surface water partial pressure of CO_2 database: Measurements performed during 1957-2013, Version 2012.

Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., and Ho, C. (2014). Climatological distributions of pH, pCO_2, total CO_2, alkalinity, and $CaCO_3$ saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chemistry*, 164:95–125.

Takahashi, T., Sutherland, S. C., and Kozyr, A. (2017). Global Ocean Surface Water Partial Pressure of CO2 Database: Measurements Performed During 1957-2016 (LDEO Database Version 2016) Version 3.3.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. a., Sabine, C., Olafsson, J., and Nojiri, Y. (2002). Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(9-10):1601–1622.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W. (2009). Climatological mean and decadal change in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 56(8-10):554–577.

Talley, L., Feely, R., Sloyan, B., Wanninkhof, R., Baringer, M., Bullister, J., Carlson, C., Doney, S., Fine, R., Firing, E., Gruber, N., Hansell, D., Ishii, M., Johnson, G., Katsumata, K., Key, R., Kramp, M., Langdon, C., Macdonald, A., Mathis, J., McDonagh, E., Mecking, S., Millero, F., Mordy, C., Nakano, T., Sabine, C., Smethie, W., Swift, J., Tanhua, T., Thurnherr, A., Warner, M., and Zhang, J.-Z. (2016). Changes in Ocean Heat, Carbon Content, and Ventilation: A Review of the First Decade of GO-SHIP Global Repeat Hydrography. *Annual Review of Marine Science*, 8(1):185–215.

Talley, L., Pickard, G., Emery, W., and Swift, J. (2011). *Descriptive Physical Oceanography: An Introduction*. Elsevier.

Tanhua, T., van Heuven, S., Key, R. M., Velo, A., Olsen, A., Schirnick, C., T. Tanhua, and Heuven, S. V. (2010). Quality control procedures and methods of the CARINA database. *Earth System Science Data*, 2(1):35–49.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106:7183–7192.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *American Meteorological Society*, 93:485–498.

Thoning, K., Kitzis, D., and Crotwell, A. (2016). Atmospheric Carbon Dioxide Dry Air Mole Fractions from quasi-continuous measurements at Mauna Loa, Hawaii. Annually updated. National Oceanic and Atmospheric Administration (NOAA), Earth System Research Laboratory (ESRL), Global Monitoring Division (GMD): Boulder, Colorado, USA. Version 2017-8.

Tjiputra, J. F., Olsen, A., Bopp, L., Lenton, A., Pfeil, B., Roy, T., Segschneider, J., Totterdell, I., and Heinze, C. (2014). Long-term surface pCO_2 trends from observations and models. *Tellus, Series B: Chemical and Physical Meteorology*, 66(1).

Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., Seland, O., and Heinze, C. (2013). Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geoscientific Model Development Discussions*, 5(4):3035–3087.

Toggweiler, J. R., Russell, J. L., and Carson, S. R. (2006). Midlatitude westerlies, atmospheric CO_2, and climate change

during the ice ages. *Paleoceanography*, 21:PA2005.

Ullman, D. J., McKinley, G. A., Bennington, V., and Dutkiewicz, S. (2009). Trends in the North Atlantic carbon sink: 1992-2006. *Global Biogeochemical Cycles*, 23(4):1992–2006.

Valsala, V. and Maksyutov, S. (2010). Simulation and assimilation of global ocean pCO_2 and air-sea CO_2 fluxes using ship observations of surface ocean pCO_2 in a simplified biogeochemical offline model. *Tellus B: Chemical and Physical Meteorology*, 62B:821–840.

Van Heuven, S., Pierrot, D., Rae, J., Lewis, E., and Wallace, D. (2011). CO2SYS v 1.1: MATLAB Program Developed for CO_2 System Calculations. ORNL/CDIAC-105b.

Van Vuuren, D., Den Elzen, M., Lucas, P., Eickhout, B., Strengers, B., Van Ruijven, B., Wonink, S., and Van Houdt, R. (2007). Stabilizing greenhouse gas concentrations at low levels: An assessment of reduction strategies and costs. *Climatic Change*, 81:119–159.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, 109:5–31.

Vichi, M., Manzini, E., Fogli, P., Alessandri, A., Patara, L., Scoccimarro, E., Masina, S., and Navarra, A. (2011). Global and regional ocean carbon uptake and climate change: Sensitivity to an aggressive mitigation scenario. *Climate Dynamics*, 37:1929–1947.

Vichi, M., Pinardi, N., and S., M. (2007). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *Journal of Marine Systems*, 64:89–109.

Völker, C. (2002). On the role of heat fluxes in the uptake of anthropogenic carbon in the North Atlantic. *Global Biogeochemical Cycles*, 16(4):1–9.

Volodin, E. M., Dianskii, N. A., and Gusev, A. V. (2010). Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya, Atmospheric and Oceanic Physics*, 46(4):414–431.

Wanninkhof, R. (1992). Relationship Between Wind Speed and Gas Ex-

change. *Journal of Geophysical Research*, 97(92):7373–7382.

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, 12(6):351–362.

Wanninkhof, R., Park, G. H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., Gruber, N., Doney, S. C., McKinley, G. A., Lenton, A., Le Quéré, C., Heinze, C., Schwinger, J., Graven, H., and Khatiwala, S. (2013). Global ocean carbon uptake: magnitude, variability and trends. *Biogeosciences*, 10(3):1983–2000.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M. (2011). MIROC-ESM: model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development Discussions*, 4(2):1063–1128.

Watson, A. J., Schuster, U., Bakker, D. C. E., Bates, N. R., Corbière, A., González-Dávila, M., Friedrich, T., Hauck, J., Heinze, C., Johannessen, T., Körtzinger, A., Metzl, N., Olafsson, J., Olsen, A., Oschlies, A., Padin, X. A., Pfeil, B., Santana-Casiano, J. M., Steinhoff, T., Telszewski, M., Rios, A. F., Wallace, D. W. R., and Wanninkhof, R. (2009). Tracking the variable North Atlantic sink for atmospheric CO_2. *Science (New York, N.Y.)*, 326(5958):1391–1393.

Waugh, D. W., Hall, T. M., McNeil, B. I., Key, R., and Matear, R. J. (2006). Anthropogenic CO_2 in the oceans estimated using transit time distributions. *Tellus*, (58B):376–389.

Weaver, A. J., Eby, M., Kienast, M., and Saenko, O. A. (2007). Response of the Atlantic meridional overturning circulation to increasing atmospheric CO_2: Sensitivity to mean climate state. *Geophysical Research Letters*, 34(5):1–5.

Webster, R. and Oliver, M. (2007). *Geostatistics for Environmental Scientists*, volume 2nd Edition.

Weiss, R. (1974). Carbon dioxide in water and seawater; the solubility of a non-ideal gas. *Marine Chemistry*, 2:203–215.

While, J., Totterdell, I., and Martin, M. (2012). Assimilation of pCO_2 data into a global coupled physical-biogeochemical ocean model. *Journal of Geophysical Research: Oceans*, 117(3):1–11.

Williams, R. and Follows, M. J. (2011).

Ocean Dynamics and the Carbon Cycle. Cambridge University Press.

WOCE (2002). WOCE observations 1990-1998; a summary of the WOCE global data resources. Technical Report 179/02, Southampton, UK.

Wolf-Gladrow, D., Zeebe, R. E., Klaas, C., Körtzinger, A., and Dickson, A. G. (2007). Total Alkalinity: The explicit conservative expression and its application to biogeochemical processes. *Marine Chemistry*, 106:287–300.

Wu, T., Li, W., Ji, J., Xin, X., Li, L., Wang, Z., Zhang, Y., Li, J., Zhang, F., Wei, M., Shi, X., Wu, F., Zhang, L., Chu, M., Jie, W., Liu, Y., Wang, F., Liu, X., Li, Q., Dong, M., Liang, X., Gao, Y., and Zhang, J. (2013). Global carbon budgets simulated by the beijing climate center climate system model for the last century. *Journal of Geophysical Research: Atmospheres*, 118:4326–4347.

Wu, T., Song, L., Li, W., Wang, Z., Zhang, H., Xin, X., Zhang, Y., Zhang, L., Li, J., Wu, F., Liu, Y., Zhang, F., Shi, X., Chu, M., Zhang, J., Fang, Y., Wang, F., Lu, Y., Liu, X., Wei, M., Liu, Q., Zhou, W., Dong, M., Zhao, Q., Ji, J., Li, L., and Zhou, M. (2008). An overview of BCC climate system model development and application for climate change studies. *Journal of Meteorological Research*, 28:034–056.

Xie, S.-P. and Kosaka, Y. (2017). What caused the global surface warming hiatus of 1998-2013? *Current Climate Change Reports*, 3:128–140.

Yashayaev, I., Bersch, M., and van Aken, H. M. (2007). Spreading of the Labrador Sea Water to the Irminger and Iceland basins. *Geophysical Research Letters*, 34(L10602).

Yashayaev, I. and Loder, J. W. (2016). Further intensification of deep convection in the Labrador Sea in 2016. *Geophysical Research Letters*, 44:1429–1438.

Yukimoto, S., Yoshimura, H., Hosaka, M., Sakami, T., Tsujino, H., Hirabara, M., Tanaka, T. Y., Deushi, M., Obata, A., Nakano, H., Y., A., Shindo, E., S., Y., Ose, T., and Kitoh, A. (2011). Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1), Model Description. Technical report, Meteorological Research Institute, Japan.

Zahariev, K., Christian, J., and Denman, K. (2008). Preindustrial, historical, and fertilization simulations using a global ocean carbon model with new parameterizations of iron limitation, calcification, and

N_2 fixation. *Progress in Oceanography*, 77(1):56–82.

Zeebe, R. E. and Wolf-Gladrow, D. (2001). *CO_2 in Seawater: Equilibrium, Kinetics, Isotopes*. Elsevier.